(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 July 2001 (19.07.2001)

PCT

(10) International Publication Number
WO 01/51638 A2

(51) International Patent Classification[7]: C12N 15/52, 9/00, 15/63, 5/10, A01K 67/027, C12P 21/00, C07K 16/40, C12Q 1/68, A61K 38/43, C12Q 1/00, G01N 33/573

(21) International Application Number: PCT/US01/01174

(22) International Filing Date: 12 January 2001 (12.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/176,139	14 January 2000 (14.01.2000)	US
60/177,443	21 January 2000 (21.01.2000)	US
60/178,574	28 January 2000 (28.01.2000)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **YANG, Junming** [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **AU-YOUNG, Janice** [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **REDDY, Roopa** [IN/US]; 1233 W. McKinley Avenue #3, Sunnyvale, CA 94086 (US). **RING, Huijun, Z.** [US/US]; 625 Orange Avenue, Los Altos, CA 94022 (US). **HILLMAN, Jennifer, L.** [US/US]; 230 Monroe Drive #17, Mountain View, CA 94040 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **YAO, Monique, G.** [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). **GANDHI, Ameena, R.** [US/US]; 837 Roble Avenue #1, Menlo Park, CA 94025 (US). **NGUYEN, Danniel, B.** [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). **TANG, Y., Tom** [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **LAL, Preeti** [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). **BANDMAN, Olga** [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/51638 A2

(54) Title: DRUG METABOLIZING ENZYMES

(57) Abstract: The invention provides human drug metabolizing enzymes (DME) and polynucleotides which identify and encode DME. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of DME.

DRUG METABOLIZING ENZYMES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of drug metabolizing enzymes and to the use of these sequences in the diagnosis, treatment, and prevention of autoimmune/inflammatory, cell proliferative, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of drug metabolizing enzymes.

BACKGROUND OF THE INVENTION

The metabolism of a drug and its movement through the body (pharmacokinetics) are important in determining its effects, toxicity, and interactions with other drugs. The three processes governing pharmacokinetics are the absorption of the drug, distribution to various tissues, and elimination of drug metabolites. These processes are intimately coupled to drug metabolism, since a variety of metabolic modifications alter most of the physicochemical and pharmacological properties of drugs, including solubility, binding to receptors, and excretion rates. The metabolic pathways which modify drugs also accept a variety of naturally occurring substrates such as steroids, fatty acids, prostaglandins, leukotrienes, and vitamins. The enzymes in these pathways are therefore important sites of biochemical and pharmacological interaction between natural compounds, drugs, carcinogens, mutagens, and xenobiotics.

It has long been appreciated that inherited differences in drug metabolism lead to drastically different levels of drug efficacy and toxicity among individuals. For drugs with narrow therapeutic indices, or drugs which require bioactivation (such as codeine), these polymorphisms can be critical. Moreover, promising new drugs are frequently eliminated in clinical trials based on toxicities which may only affect a segment of the patient group. Advances in pharmacogenomics research, of which drug metabolizing enzymes constitute an important part, are promising to expand the tools and information that can be brought to bear on questions of drug efficacy and toxicity (See Evans, W. E. and R. V. Relling (1999) Science 286:487-491).

Drug metabolic reactions are categorized as Phase I, which functionalize the drug molecule and prepare it for further metabolism, and Phase II, which are conjugative. In general, Phase I reaction products are partially or fully inactive, and Phase II reaction products are the chief excreted species. However, Phase I reaction products are sometimes more active than the original administered drugs; this metabolic activation principle is exploited by pro-drugs (e.g. L-dopa). Additionally, some nontoxic compounds (e.g. aflatoxin, benzo[*a*]pyrene) are metabolized to toxic

intermediates through these pathways. Phase I reactions are usually rate-limiting in drug metabolism. Prior exposure to the compound, or other compounds, can induce the expression of Phase I enzymes however, and thereby increase substrate flux through the metabolic pathways. (See Klaassen, C. D., Amdur, M. O. and J. Doull (1996) Casarett and Doull's Toxicology: The Basic Science of Poisons, McGraw-Hill, New York, NY, pp. 113-186; B. G. Katzung (1995) Basic and Clinical Pharmacology, Appleton and Lange, Norwalk, CT, pp. 48-59; G. G. Gibson and P. Skett (1994) Introduction to Drug Metabolism, Blackie Academic and Professional, London.)

Drug metabolizing enzymes (DMEs) have broad substrate specificities. This can be contrasted to the immune system, where a large and diverse population of antibodies are highly specific for their antigens. The ability of DMEs to metabolize a wide variety of molecules creates the potential for drug interactions at the level of metabolism. For example, the induction of a DME by one compound may affect the metabolism of another compound by the enzyme.

DMEs have been classified according to the type of reaction they catalyze and the cofactors involved. The major classes of Phase I enzymes include, but are not limited to, cytochrome P450 and flavin-containing monooxygenase. Other enzyme classes involved in Phase I-type catalytic cycles and reactions include, but are not limited to, NADPH cytochrome P450 reductase (CPR), the microsomal cytochrome b5/NADH cytochrome b5 reductase system, the ferredoxin/ferredoxin reductase redox pair, aldo/keto reductases, and alcohol dehydrogenases. The major classes of Phase II enzymes include, but are not limited to, UDP glucuronyltransferase, sulfotransferase, glutathione S-transferase, N-acyltransferase, and N-acetyl transferase.

Cytochrome P450 and P450 catalytic cycle-associated enzymes

Members of the cytochrome P450 superfamily of enzymes catalyze the oxidative metabolism of a variety of substrates, including natural compounds such as steroids, fatty acids, prostaglandins, leukotrienes, and vitamins, as well as drugs, carcinogens, mutagens, and xenobiotics. Cytochromes P450, also known as P450 heme-thiolate proteins, usually act as terminal oxidases in multi-component electron transfer chains, called P450-containing monooxygenase systems. Specific reactions catalyzed include hydroxylation, epoxidation, N-oxidation, sulfooxidation, N-, S-, and O-dealkylations, desulfation, deamination, and reduction of azo, nitro, and N-oxide groups. These reactions are involved in steroidogenesis of glucocorticoids, cortisols, estrogens, and androgens in animals; insecticide resistance in insects; herbicide resistance and flower coloring in plants; and environmental bioremediation by microorganisms. Cytochrome P450 actions on drugs, carcinogens, mutagens, and xenobiotics can result in detoxification or in conversion of the substance to a more toxic product. Cytochromes P450 are abundant in the liver, but also occur in other tissues; the enzymes are located in microsomes. (See ExPASY ENZYME EC 1.14.14.1; Prosite PDOC00081 Cytochrome P450 cysteine heme-iron ligand signature; PRINTS EP450I E-Class P450 Group I

signature; Graham-Lorence, S. and Peterson, J.A. (1996) FASEB J. 10:206-214.)

Four hundred cytochromes P450 have been identified in diverse organisms including bacteria, fungi, plants, and animals (Graham-Lorence, supra). The B-class is found in prokaryotes and fungi, while the E-class is found in bacteria, plants, insects, vertebrates, and mammals. Five subclasses or groups are found within the larger family of E-class cytochromes P450 (PRINTS EP450I E-Class P450 Group I signature).

All cytochromes P450 use a heme cofactor and share structural attributes. Most cytochromes P450 are 400 to 530 amino acids in length. The secondary structure of the enzyme is about 70% alpha-helical and about 22% beta-sheet. The region around the heme-binding site in the C-terminal part of the protein is conserved among cytochromes P450. A ten amino acid signature sequence in this heme-iron ligand region has been identified which includes a conserved cysteine involved in binding the heme iron in the fifth coordination site. In eukaryotic cytochromes P450, a membrane-spanning region is usually found in the first 15-20 amino acids of the protein, generally consisting of approximately 15 hydrophobic residues followed by a positively charged residue. (See Prosite PDOC00081, supra; Graham-Lorence, supra.)

Cytochrome P450 enzymes are involved in cell proliferation and development. The enzymes have roles in chemical mutagenesis and carcinogenesis by metabolizing chemicals to reactive intermediates that form adducts with DNA (Nebert, D.W. and Gonzalez, F.J. (1987) Ann. Rev. Biochem. 56:945-993). These adducts can cause nucleotide changes and DNA rearrangements that lead to oncogenesis. Cytochrome P450 expression in liver and other tissues is induced by xenobiotics such as polycyclic aromatic hydrocarbons, peroxisomal proliferators, phenobarbital, and the glucocorticoid dexamethasone (Dogra, S.C. et al. (1998) Clin. Exp. Pharmacol. Physiol. 25:1-9). A cytochrome P450 protein may participate in eye development as mutations in the P450 gene CYP1B1 cause primary congenital glaucoma (Online Mendelian Inheritance in Man (OMIM) *601771 Cytochrome P450, subfamily I (dioxin-inducible), polypeptide 1; CYP1B1).

Cytochromes P450 are associated with inflammation and infection. Hepatic cytochrome P450 activities are profoundly affected by various infections and inflammatory stimuli, some of which are suppressed and some induced (Morgan, E.T. (1997) Drug Metab. Rev. 29:1129-1188). Effects observed in vivo can be mimicked by proinflammatory cytokines and interferons. Autoantibodies to two cytochrome P450 proteins were found in patients with autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy (APECED), a polyglandular autoimmune syndrome (OMIM *240300 Autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy).

Mutations in cytochromes P450 have been linked to metabolic disorders, including congenital adrenal hyperplasia, the most common adrenal disorder of infancy and childhood; pseudovitamin D-deficiency rickets; cerebrotendinous xanthomatosis, a lipid storage disease characterized by

progressive neurologic dysfunction, premature atherosclerosis, and cataracts; and an inherited resistance to the anticoagulant drugs coumarin and warfarin (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, Inc. New York, NY, pp. 1968-1970; Takeyama, K. et al. (1997) Science 277:1827-1830; Kitanaka, S. et al. (1998) N. Engl. J. Med. 338:653-661; OMIM *213700 Cerebrotendinous xanthomatosis; and OMIM #122700 Coumarin resistance). Extremely high levels of expression of the cytochrome P450 protein aromatase were found in a fibrolamellar hepatocellular carcinoma from a boy with severe gynecomastia (feminization) (Agarwal, V.R. (1998) J. Clin. Endocrinol. Metab. 83:1797-1800).

The cytochrome P450 catalytic cycle is completed through reduction of cytochrome P450 by NADPH cytochrome P450 reductase (CPR). Another microsomal electron transport system consisting of cytochrome b5 and NADPH cytochrome b5 reductase has been widely viewed as a minor contributor of electrons to the cytochrome P450 catalytic cycle. However, a recent report by Lamb, D. C. et al. (1999; FEBS Lett. 462:283-8) identifies a Candida albicans cytochrome P450 (CYP51) which can be efficiently reduced and supported by the microsomal cytochrome b5/NADPH cytochrome b5 reductase system. Therefore, there are likely many cytochromes P450 which are supported by this alternative electron donor system.

Cytochrome b5 reductase is also responsible for the reduction of oxidized hemoglobin (methemoglobin, or ferrihemoglobin, which is unable to carry oxygen) to the active hemoglobin (ferrohemoglobin) in red blood cells. Methemoglobinemia results when there is a high level of oxidant drugs or an abnormal hemoglobin (hemoglobin M) which is not efficiently reduced. Methemoglobinemia can also result from a hereditary deficiency in red cell cytochrome b5 reductase (Reviewed in Mansour, A. and Lurie, A. A. (1993) Am. J. Hematol. 42:7-12).

Members of the cytochrome P450 family are also closely associated with vitamin D synthesis and catabolism. Vitamin D exists as two biologically equivalent prohormones, ergocalciferol (vitamin D_2), produced in plant tissues, and cholecalciferol (vitamin D_3), produced in animal tissues. The latter form, cholecalciferol, is formed upon the exposure of 7-dehydrocholesterol to near ultraviolet light (i.e., 290-310 nm), normally resulting from even minimal periods of skin exposure to sunlight (reviewed in Miller, W.L. and Portale, A.A. (2000) Trends in Endocrinology and Metabolism 11:315-319).

Both prohormone forms are further metabolized in the liver to 25-hydroxyvitamin D (25(OH)D) by the enzyme 25-hydroxylase. 25(OH)D is the most abundant precursor form of vitamin D which must be further metabolized in the kidney to the active form, 1α,25-dihydroxyvitamin D (1α,25$(OH)_2$D), by the enzyme 25-hydroxyvitamin D 1α-hydroxylase (1α-hydroxylase). Regulation of 1α,25$(OH)_2$D production is primarily at this final step in the synthetic pathway. The activity of 1α-hydroxylase depends upon several physiological factors including the circulating level of the

enzyme product ($1\alpha,25(OH)_2D$) and the levels of parathyroid hormone (PTH), calcitonin, insulin, calcium, phosphorus, growth hormone, and prolactin. Furthermore, extrarenal 1α-hydroxylase activity has been reported, suggesting that tissue-specific, local regulation of 1α,25(OH)2D production may also be biologically important. The catalysis of $1\alpha,25(OH)_2D$ to 24,25-dihydroxyvitamin D ($24,25(OH)_2D$), involving the enzyme 25-hydroxyvitamin D 24-hydroxylase (24-hydroxylase), also occurs in the kidney. 24-hydroxylase can also use 25(OH)D as a substrate (Shinki, T. et al. (1997) Proc. Natl. Acad. Sci. U.S.A. 94:12920-12925; Miller, W.L. and Portale, A.A. supra; and references within).

Vitamin D 25-hydroxylase, 1α-hydroxylase, and 24-hydroxylase are all NADPH-dependent, type I (mitochondrial) cytochrome P450 enzymes that show a high degree of homology with other members of the family. Vitamin D 25-hydroxylase also shows a broad substrate specificity and may also perform 26-hydroxylation of bile acid intermediates and 25, 26, and 27-hydroxylation of cholesterol (Dilworth, F.J. et al. (1995) J. Biol. Chem. 270:16766-16774; Miller, W.L. and Portale, A.A. supra; and references within).

The active form of vitamin D ($1\alpha,25(OH)_2D$) is involved in calcium and phosphate homeostasis and promotes the differentiation of myeloid and skin cells. Vitamin D deficiency resulting from deficiencies in the enzymes involved in vitamin D metabolism (e.g., 1α-hydroxylase) causes hypocalcemia, hypophosphatemia, and vitamin D-dependent (sensitive) rickets, a disease characterized by loss of bone density and distinctive clinical features, including bandy or bow leggedness accompanied by a waddling gait. Deficiencies in vitamin D 25-hydroxylase cause cerebrotendinous xanthomatosis, a lipid-storage disease characterized by the deposition of cholesterol and cholestanol in the Achilles' tendons, brain, lungs, and many other tissues. The disease presents with progressive neurologic dysfunction, including postpubescent cerebellar ataxia, atherosclerosis, and cataracts. Vitamin D 25-hydroxylase deficiency does not result in rickets, suggesting the existence of alternative pathways for the synthesis of 25(OH)D (Griffin, J.E. and Zerwekh, J.E. (1983) J. Clin. Invest. 72:1190-1199; Gamblin, G.T. et al. (1985) J. Clin. Invest. 75:954-960; and W.L. and Portale, A.A. supra).

Ferredoxin and ferredoxin reductase are electron transport accessory proteins which support at least one human cytochrome P450 species, cytochrome P450c27 encoded by the CYP27 gene (Dilworth, F. J. et al. (1996) Biochem. J. 320:267-71). A Streptomyces griseus cytochrome P450, CYP104D1, was heterologously expressed in E. coli and found to be reduced by the endogenous ferredoxin and ferredoxin reductase enzymes (Taylor, M. et al. (1999) Biochem. Biophys. Res. Commun. 263:838-42), suggesting that many cytochrome P450 species may be supported by the ferredoxin/ferredoxin reductase pair. Ferredoxin reductase has also been found in a model drug metabolism system to reduce actinomycin D, an antitumor antibiotic, to a reactive free radical species

(Flitter, W. D. and Mason, R. P. (1988) Arch. Biochem. Biophys. 267:632-9).

Flavin-containing monooxygenase (FMO)

Flavin-containing monooxygenases oxidize the nucleophilic nitrogen, sulfur, and phosphorus heteroatom of an exceptional range of substrates. Like cytochromes P450, FMOs are microsomal and use NADPH and O_2; there is also a great deal of substrate overlap with cytochromes P450. The tissue distribution of FMOs includes liver, kidney, and lung.

There are five different known isoforms of FMO in mammals (FMO1, FMO2, FMO3, FMO4, and FMO5), which are expressed in a tissue-specific manner. The isoforms differ in their substrate specificities and other properties such as inhibition by various compounds and stereospecificity of reaction. FMOs have a 13 amino acid signature sequence, the components of which span the N-terminal two-thirds of the sequences and include the FAD binding region and the FATGY motif which has been found in many N-hydroxylating enzymes (Stehr, M. et al. (1998) Trends Biochem. Sci. 23:56-57; PRINTS FMOXYGENASE Flavin-containing monooxygenase signature).

Specific reactions include oxidation of nucleophilic tertiary amines to N-oxides, secondary amines to hydroxylamines and nitrones, primary amines to hydroxylamines and oximes, and sulfur-containing compounds and phosphines to S- and P-oxides. Hydrazines, iodides, selenides, and boron-containing compounds are also substrates. Although FMOs appear similar to cytochromes P450 in their chemistry, they can generally be distinguished from cytochromes P450 in vitro based on, for example, the higher heat lability of FMOs and the nonionic detergent sensitivity of cytochromes P450; however, use of these properties in identification is complicated by further variation among FMO isoforms with respect to thermal stability and detergent sensitivity.

FMOs play important roles in the metabolism of several drugs and xenobiotics. FMO (FMO3 in liver) is predominantly responsible for metabolizing (S)-nicotine to (S)-nicotine N-1'-oxide, which is excreted in urine. FMO is also involved in S-oxygenation of cimetidine, an H_2-antagonist widely used for the treatment of gastric ulcers. Liver-expressed forms of FMO are not under the same regulatory control as cytochrome P450. In rats, for example, phenobarbital treatment leads to the induction of cytochrome P450, but the repression of FMO1.

Endogenous substrates of FMO include cysteamine, which is oxidized to the disulfide, cystamine, and trimethylamine (TMA), which is metabolized to trimethylamine N-oxide. TMA smells like rotting fish, and mutations in the FMO3 isoform lead to large amounts of the malodorous free amine being excreted in sweat, urine, and breath. These symptoms have led to the designation fish-odor syndrome (OMIM 602079 Trimethylaminuria).

Lysyl oxidase:

Lysyl oxidase (lysine 6-oxidase, LO) is a copper-dependent amine oxidase involved in the

formation of connective tissue matrices by crosslinking collagen and elastin. LO is secreted as a N-glycosylated precuror protein of approximately 50 kDa Levels and cleaved to the mature form of the enzyme by a metalloprotease, although the precursor form is also active. The copper atom in LO is involved in the transport of electron to and from oxygen to facilitate the oxidative deamination of lysine residues in these extracellular matrix proteins. While the coordination of copper is essential to LO activity, insufficient dietary intake of copper does not influence the expression of the apoenzyme. However, the absence of the functional LO is linked to the skeletal and vascular tissue disorders that are associated with dietary copper deficiency. LO is also inhibited by a variety of semicarbazides, hydrazines, and amino nitrites, as well as heparin. Beta-aminopropionitrile is a commonly used inhibitor. LO activity is increased in response to ozone, cadmium, and elevated levels of hormones released in response to local tissue trauma, such as transforming growth factor-beta, platelet-derived growth factor, angiotensin II, and fibroblast growth factor. Abnormalities in LO activity has been linked to Menkes syndrome and occipital horn syndrome. Cytosolic forms of the enzyme hae been implicated in abnormal cell proliferation (reviewed in Rucker, R.B. et al. (1998) Am. J. Clin. Nutr. 67:996S-1002S and Smith-Mungo. L.I. and Kagan, H.M. (1998) Matrix Biol. 16:387-398).

Dihydrofolate reductases

Dihydrofolate reductases (DHFR) are ubiquitous enzymes that catalyze the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate, an essential step in the de novo synthesis of glycine and purines as well as the conversion of deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP). The basic reaction is as follows:

$$\text{7,8-dihydrofolate} + \text{NADPH} \rightarrow \text{5,6,7,8-tetrahydrofolate} + \text{NADP}^+$$

The enzymes can be inhibited by a number of dihydrofolate analogs, including trimethroprim and methotrexate. Since an abundance of TMP is required for DNA synthesis, rapidly dividing cells require the activity of DHFR. The replication of DNA viruses (i.e., herpesvirus) also requires high levels of DHFR activity. As a result, drugs that target DHFR have been used for cancer chemotherapy and to inhibit DNA virus replication. (For similar reasons, thymidylate synthetases are also target enzymes.) Drugs that inhibit DHFR are preferentially cytotoxic for rapidly dividing cells (or DNA virus-infected cells) but have no specificity, resulting in the indiscriminate destruction of dividing cells. Furthermore, cancer cells may become resistant to drugs such as methotrexate as a result of acquired transport defects or the duplication of one or more DHFR genes (Stryer, L (1988) Biochemistry. W.H Freeman and Co., Inc. New York. pp. 511-5619).

Aldo/keto reductases

Aldo/keto reductases are monomeric NADPH-dependent oxidoreductases with broad

substrate specificities (Bohren, K. M. et al. (1989) J. Biol. Chem. 264:9547-51). These enzymes catalyze the reduction of carbonyl-containing compounds, including carbonyl-containing sugars and aromatic compounds, to the corresponding alcohols. Therefore, a variety of carbonyl-containing drugs and xenobiotics are likely metabolized by enzymes of this class.

One known reaction catalyzed by a family member, aldose reductase, is the reduction of glucose to sorbitol, which is then further metabolized to fructose by sorbitol dehydrogenase. Under normal conditions, the reduction of glucose to sorbitol is a minor pathway. In hyperglycemic states, however, the accumulation of sorbitol is implicated in the development of diabetic complications (OMIM *103880 Aldo-keto reductase family 1, member B1). Members of this enzyme family are also highly expressed in some liver cancers (Cao, D. et al. (1998) J. Biol. Chem. 273:11429-35).

Alcohol dehydrogenases

Alcohol dehydrogenases (ADHs) oxidize simple alcohols to the corresponding aldehydes. ADH is a cytosolic enzyme, prefers the cofactor NAD^+, and also binds zinc ion. Liver contains the highest levels of ADH, with lower levels in kidney, lung, and the gastric mucosa.

Known ADH isoforms are dimeric proteins composed of 40 kDa subunits. There are five known gene loci which encode these subunits (a, b, g, p, c), and some of the loci have characterized allelic variants (b_1, b_2, b_3, g_1, g_2). The subunits can form homodimers and heterodimers; the subunit composition determines the specific properties of the active enzyme. The holoenzymes have therefore been categorized as Class I (subunit compositions aa, ab, ag, bg, gg), Class II (pp), and Class III (cc). Class I ADH isozymes oxidize ethanol and other small aliphatic alcohols, and are inhibited by pyrazole. Class II isozymes prefer longer chain aliphatic and aromatic alcohols, are unable to oxidize methanol, and are not inhibited by pyrazole. Class III isozymes prefer even longer chain aliphatic alcohols (five carbons and longer) and aromatic alcohols, and are not inhibited by pyrazole.

The short-chain alcohol dehydrogenases include a number of related enzymes with a variety of substrate specificities. Included in this group are the mammalian enzymes D-beta-hydroxybutyrate dehydrogenase, (R)-3-hydroxybutyrate dehydrogenase, 15-hydroxyprostaglandin dehydrogenase, NADPH-dependent carbonyl reductase, corticosteroid 11-beta-dehydrogenase, and estradiol 17-beta-dehydrogenase, as well as the bacterial enzymes acetoacetyl-CoA reductase, glucose 1-dehydrogenase, 3-beta-hydroxysteroid dehydrogenase, 20-beta-hydroxysteroid dehydrogenase, ribitol dehydrogenase, 3-oxoacyl reductase, 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase, sorbitol-6-phosphate 2-dehydrogenase, 7-alpha-hydroxysteroid dehydrogenase, cis-1,2-dihydroxy-3,4-cyclohexadiene-1-carboxylate dehydrogenase, *cis*-toluene dihydrodiol dehydrogenase, *cis*-benzene glycol dehydrogenase, biphenyl-2,3-dihydro-2,3-diol dehydrogenase, N-acylmannosamine 1-dehydrogenase, and 2-deoxy-D-gluconate 3-dehydrogenase (Krozowski, Z. (1994) J. Steroid

Biochem. Mol. Biol. 51:125-130; Krozowski, Z. (1992) Mol. Cell Endocrinol. 84:C25-31; and Marks, A.R. et al. (1992) J. Biol. Chem. 267:15459-15463).

UDP glucuronyltransferase

Members of the UDP glucuronyltransferase family (UGTs) catalyze the transfer of a glucuronic acid group from the cofactor uridine diphosphate-glucuronic acid (UDP-glucuronic acid) to a substrate. The transfer is generally to a nucleophilic heteroatom (O, N, or S). Substrates include xenobiotics which have been functionalized by Phase I reactions, as well as endogenous compounds such as bilirubin, steroid hormones, and thyroid hormones. Products of glucuronidation are excreted in urine if the molecular weight of the substrate is less than about 250 g/mol, whereas larger glucuronidated substrates are excreted in bile.

UGTs are located in the microsomes of liver, kidney, intestine, skin, brain, spleen, and nasal mucosa, where they are on the same side of the endoplasmic reticulum membrane as cytochrome P450 enzymes and flavin-containing monooxygenases, and therefore are ideally located to access products of Phase I drug metabolism. UGTs have a C-terminal membrane-spanning domain which anchors them in the endoplasmic reticulum membrane, and a conserved signature domain of about 50 amino acid residues in their C terminal section (Prosite PDOC00359 UDP-glycosyltransferase signature).

UGTs involved in drug metabolism are encoded by two gene families, UGT1 and UGT2. Members of the UGT1 family result from alternative splicing of a single gene locus, which has a variable substrate binding domain and constant region involved in cofactor binding and membrane insertion. Members of the UGT2 family are encoded by separate gene loci, and are divided into two families, UGT2A and UGT2B. The 2A subfamily is expressed in olfactory epithelium, and the 2B subfamily is expressed in liver microsomes. Mutations in UGT genes are associated with hyperbilirubinemia (OMIM #143500 Hyperbilirubinemia I); Crigler-Najjar syndrome, characterized by intense hyperbilirubinemia from birth (OMIM #218800 Crigler-Najjar syndrome); and a milder form of hyperbilirubinemia termed Gilbert's disease (OMIM *191740 UGT1).

Sulfotransferase

Sulfate conjugation occurs on many of the same substrates which undergo O-glucuronidation to produce a highly water-soluble sulfuric acid ester. Sulfotransferases (ST) catalyze this reaction by transferring SO_3^- from the cofactor 3'-phosphoadenosine-5'-phosphosulfate (PAPS) to the substrate. ST substrates are predominantly phenols and aliphatic alcohols, but also include aromatic amines and aliphatic amines, which are conjugated to produce the corresponding sulfamates. The products of these reactions are excreted mainly in urine.

STs are found in a wide range of tissues, including liver, kidney, intestinal tract, lung, platelets, and brain. The enzymes are generally cytosolic, and multiple forms are often co-expressed.

For example, there are more than a dozen forms of ST in rat liver cytosol. These biochemically characterized STs fall into five classes based on their substrate preference: arylsulfotransferase, alcohol sulfotransferase, estrogen sulfotransferase, tyrosine ester sulfotransferase, and bile salt sulfotransferase.

ST enzyme activity varies greatly with sex and age in rats. The combined effects of developmental cues and sex-related hormones are thought to lead to these differences in ST expression profiles, as well as the profiles of other DMEs such as cytochromes P450. Notably, the high expression of STs in cats partially compensates for their low level of UDP glucuronyltransferase activity.

Several forms of ST have been purified from human liver cytosol and cloned. There are two phenol sulfotransferases with different thermal stabilities and substrate preferences. The thermostable enzyme catalyzes the sulfation of phenols such as para-nitrophenol, minoxidil, and acetaminophen; the thermolabile enzyme prefers monoamine substrates such as dopamine, epinephrine, and levadopa. Other cloned STs include an estrogen sulfotransferase and an N-acetylglucosamine-6-O-sulfotransferase. This last enzyme is illustrative of the other major role of STs in cellular biochemistry, the modification of carbohydrate structures that may be important in cellular differentiation and maturation of proteoglycans. Indeed, an inherited defect in a sulfotransferase has been implicated in macular corneal dystrophy, a disorder characterized by a failure to synthesize mature keratan sulfate proteoglycans (Nakazawa, K. et al. (1984) J. Biol. Chem. 259:13751-7; OMIM *217800 Macular dystrophy, corneal).

Galactosyltransferases

Galactosyltransferases are a subset of glycosyltransferases that transfer galactose (Gal) to the terminal N-acetylglucosamine (GlcNAc) oligosaccharide chains that are part of glycoproteins or glycolipids that are free in solution (Kolbinger, F. et al. (1998) J. Biol. Chem. 273:433-440; Amado, M. et al. (1999) Biochim. Biophys. Acta 1473:35-53). Galactosyltransferases have been detected on the cell surface and as soluble extracellular proteins, in addition to being present in the Golgi. β1,3-galactosyltransferases form Type I carbohydrate chains with Gal (β1-3)GlcNAc linkages. Known human and mouse β1,3-galactosyltransferases appear to have a short cytosolic domain, a single transmembrane domain, and a catalytic domain with eight conserved regions. (Kolbinger, F. supra and Hennet, T. et al. (1998) J. Biol. Chem. 273:58-65). In mouse UDP-galactose:β-N-acetylglucosamine β1,3-galactosyltransferase-I region 1 is located at amino acid residues 78-83, region 2 is located at amino acid residues 93-102, region 3 is located at amino acid residues 116-119, region 4 is located at amino acid residues 147-158, region 5 is located at amino acid residues 172-183, region 6 is located at amino acid residues 203-206, region 7 is located at amino acid residues 236-246, and region 8 is located at amino acid residues 264-275. A variant of a sequence found

within mouse UDP-galactose:β-N-acetylglucosamine β1,3-galactosyltransferase-I region 8 is also found in bacterial galactosyltransferases, suggesting that this sequence defines a galactosyltransferase sequence motif (Hennet, T. supra). Recent work suggests that brainiac protein is a β1,3-galactosyltransferase. (Yuan, Y. et al. (1997) Cell 88:9-11; and Hennet, T. supra).

UDP-Gal:GlcNAc-1,4-galactosyltransferase (-1,4-GalT) (Sato, T. et al., (1997) EMBO J. 16:1850-1857) catalyzes the formation of Type II carbohydrate chains with Gal (β1-4)GlcNAc linkages. As is the case with the β1,3-galactosyltransferase, a soluble form of the enzyme is formed by cleavage of the membrane-bound form. Amino acids conserved among β1,4-galactosyltransferases include two cysteines linked through a disulfide-bonded and a putative UDP-galactose-binding site in the catalytic domain (Yadav, S. and Brew, K. (1990) J. Biol. Chem. 265:14163-14169; Yadav, S.P. and Brew, K. (1991) J. Biol. Chem. 266:698-703; and Shaper, N.L. et al. (1997) J. Biol. Chem. 272:31389-31399). β1,4-galactosyltransferases have several specialized roles in addition to synthesizing carbohydrate chains on glycoproteins or glycolipids. In mammals a β1,4-galactosyltransferase, as part of a heterodimer with α-lactalbumin, functions in lactating mammary gland lactose production. A β1,4-galactosyltransferase on the surface of sperm functions as a receptor that specifically recognizes the egg. Cell surface β1,4-galactosyltransferases also function in cell adhesion, cell/basal lamina interaction, and normal and metastatic cell migration. (Shur, B. (1993) Curr. Opin. Cell Biol. 5:854-863; and Shaper, J. (1995) Adv. Exp. Med. Biol. 376:95-104).

Glutathione S-transferase

The basic reaction catalyzed by glutathione S-transferases (GST) is the conjugation of an electrophile with reduced glutathione (GSH). GSTs are homodimeric or heterodimeric proteins localized mainly in the cytosol, but some level of activity is present in microsomes as well. The major isozymes share common structural and catalytic properties; in humans they have been classified into four major classes, Alpha, Mu, Pi, and Theta. The two largest classes, Alpha and Mu, are identified by their respective protein isoelectric points; pI ~ 7.5-9.0 (Alpha), and pI ~ 6.6 (Mu). Each GST possesses a common binding site for GSH and a variable hydrophobic binding site. The hydrophobic binding site in each isozyme is specific for particular electrophilic substrates. Specific amino acid residues within GSTs have been identified as important for these binding sites and for catalytic activity. Residues Q67, T68, D101, E104, and R131 are important for the binding of GSH (Lee, H-C et al. (1995) J. Biol. Chem. 270: 99-109). Residues R13, R20, and R69 are important for the catalytic activity of GST (Stenberg G et al. (1991) Biochem. J. 274: 549-55).

In most cases, GSTs perform the beneficial function of deactivation and detoxification of potentially mutagenic and carcinogenic chemicals. However, in some cases their action is detrimental and results in activation of chemicals with consequent mutagenic and carcinogenic

effects. Some forms of rat and human GSTs are reliable preneoplastic markers that aid in the detection of carcinogenesis. Expression of human GSTs in bacterial strains, such as Salmonella typhimurium used in the well-known Ames test for mutagenicity, has helped to establish the role of these enzymes in mutagenesis. Dihalomethanes, which produce liver tumors in mice, are believed to be activated by GST. This view is supported by the finding that dihalomethanes are more mutagenic in bacterial cells expressing human GST than in untransfected cells (Thier, R. et al. (1993) Proc. Natl. Acad. Sci. USA 90: 8567-80). The mutagenicity of ethylene dibromide and ethylene dichloride is increased in bacterial cells expressing the human Alpha GST, A1-1, while the mutagenicity of aflatoxin B1 is substantially reduced by enhancing the expression of GST (Simula, T.P. et al. (1993) Carcinogenesis 14: 1371-6). Thus, control of GST activity may be useful in the control of mutagenesis and carcinogenesis.

GST has been implicated in the acquired resistance of many cancers to drug treatment, the phenomenon known as multi-drug resistance (MDR). MDR occurs when a cancer patient is treated with a cytotoxic drug such as cyclophosphamide and subsequently becomes resistant to this drug and to a variety of other cytotoxic agents as well. Increased GST levels are associated with some of these drug resistant cancers, and it is believed that this increase occurs in response to the drug agent which is then deactivated by the GST catalyzed GSH conjugation reaction. The increased GST levels then protect the cancer cells from other cytotoxic agents which bind to GST. Increased levels of A1-1 in tumors has been linked to drug resistance induced by cyclophosphamide treatment (Dirven H.A. et al. (1994) Cancer Res. 54: 6215-20). Thus control of GST activity in cancerous tissues may be useful in treating MDR in cancer patients.

Gamma-glutamyl transpeptidase

Gamma-glutamyl transpeptidases are ubiquitously expressed enzymes that initiate extracellular glutathione (GSH) breakdown by cleaving gamma-glutamyl amide bonds. The breakdown of GSH provides cells with a regional cysteine pool for biosynthetic pathways. Gamma-glutamyl transpeptidases also contribute to cellular antioxidant defenses and expression is induced by oxidative steress. The cell surface-localized glycoproteins.are expressed at high levels in cancer cells. Studies have suggested that the high level of gamma-glutamyl transpeptidases activity present on the surface of cancer cells could be exploited to activate precursor drugs, resulting in high local concentrations of anti-cancer therapeutic agents (Hanigan, M.H. (1998) Chem. Biol. Interact. 111-112:333-42; Taniguchi, N. and Ikeda, Y. (1998) Adv. Enzymol. Relat. Areas Mol. Biol. 72:239-78; Chikhi, N. et al. (1999) Comp. Biochem. Physiol. B. Biochem. Mol. Biol. 122:367-80).

Acyltransferase

N-acyltransferase enzymes catalyze the transfer of an amino acid conjugate to an activated carboxylic group. Endogenous compounds and xenobiotics are activated by acyl-CoA synthetases in

the cytosol, microsomes, and mitochondria. The acyl-CoA intermediates are then conjugated with an amino acid (typically glycine, glutamine, or taurine, but also ornithine, arginine, histidine, serine, aspartic acid, and several dipeptides) by N-acyltransferases in the cytosol or mitochondria to form a metabolite with an amide bond. This reaction is complementary to O-glucuronidation, but amino acid conjugation does not produce the reactive and toxic metabolites which often result from glucuronidation.

One well-characterized enzyme of this class is the bile acid-CoA:amino acid N-acyltransferase (BAT) responsible for generating the bile acid conjugates which serve as detergents in the gastrointestinal tract (Falany, C. N. et al. (1994) J. Biol. Chem. 269:19375-9; Johnson, M. R. et al. (1991) J. Biol. Chem. 266:10227-33). BAT is also useful as a predictive indicator for prognosis of hepatocellular carcinoma patients after partial hepatectomy (Furutani, M. et al. (1996) Hepatology 24:1441-5).

Acetyltransferases

Acetyltransferases have been extensively studied for their role in histone acetylation. Histone acetylation results in the relaxing of the chromatin structure in eukaryotic cells, allowing transcription factors to gain access to promoter elements of the DNA templates in the affected region of the genome (or the genome in general). In contrast, histone deacetylation results in a reduction in transcription by closing the chromatin structure and limiting access of transcription factors. To this end, a common means of stimulating cell transcription is the use of chemical agents that inhibit the deacetylation of histones (e.g., sodium butyrate), resulting in a global (albeit artifactual) increase in gene expression. The modulation of gene expression by acetylation also results from the acetylation of other proteins, including but not limited to, p53, GATA-1, MyoD, ACTR, TFIIE, TFIIF and the high mobility group proteins (HMG). In the case of p53, acetylation results in increased DNA binding, leading to the stimulation of transcription of genes regulated by p53. The prototypic histone acetylase (HAT) is Gcn5 from Saccharomyces cerevisiae. Gcn5 is a member of a family of acetylases that includes Tetrahymena p55, human Gcn5, and human p300/CBP. Histone acetylation is reviewed in (Cheung, W.L. et al. (2000) Current Opinion in Cell Biology 12:326-333 and Berger, S.L (1999) Current Opinion in Cell Biology 11:336-341). Some acetyltransferase enzymes posses the alpha/beta hydrolase fold (Center of Applied Molecular Engineering Inst. of Chemistry and Biochemistry - University of Salzburg, http://predict.sanger.ac.uk/irbm-course97/Docs/ms/) common to several other major classes of enzymes, including but not limited to, acetylcholinesterases and carboxylesterases (Structural Classification of Proteins, http://scop.mrc-lmb.cam.ac.uk/scop/index.html).

N-acetyltransferase

Aromatic amines and hydrazine-containing compounds are subject to N-acetylation by the N-acetyltransferase enzymes of liver and other tissues. Some xenobiotics can be O-acetylated to some

extent by the same enzymes. N-acetyltransferases are cytosolic enzymes which utilize the cofactor acetyl-coenzyme A (acetyl-CoA) to transfer the acetyl group in a two step process. In the first step, the acetyl group is transferred from acetyl-CoA to an active site cysteine residue; in the second step, the acetyl group is transferred to the substrate amino group and the enzyme is regenerated.

In contrast to most other DME classes, there are a limited number of known N-acetyltransferases. In humans, there are two highly similar enzymes, NAT1 and NAT2; mice appear to have a third form of the enzyme, NAT3. The human forms of N-acetyltransferase have independent regulation (NAT1 is widely-expressed, whereas NAT2 is in liver and gut only) and overlapping substrate preferences. Both enzymes appear to accept most substrates to some extent, but NAT1 does prefer some substrates (para-aminobenzoic acid, para-aminosalicylic acid, sulfamethoxazole, and sulfanilamide), while NAT2 prefers others (isoniazid, hydralazine, procainamide, dapsone, aminoglutethimide, and sulfamethazine).

Clinical observations of patients taking the antituberculosis drug isoniazid in the 1950s led to the description of fast and slow acetylators of the compound. These phenotypes were shown subsequently to be due to mutations in the NAT2 gene which affected enzyme activity or stability. The slow isoniazid acetylator phenotype is very prevalent in Middle Eastern populations (approx. 70%), and is less prevalent in Caucasian (approx. 50%) and Asian (<25%) populations. More recently, functional polymorphism in NAT1 has been detected, with approximately 8% of the population tested showing a slow acetylator phenotype (Butcher, N. J. et al. (1998) Pharmacogenetics 8:67-72). Since NAT1 can activate some known aromatic amine carcinogens, polymorphism in the widely-expressed NAT1 enzyme may be important in determining cancer risk (OMIM *108345 N-acetyltransferase 1).

Aminotransferases

Aminotransferases comprise a family of pyridoxal 5'-phosphate (PLP) -dependent enzymes that catalyze transformations of amino acids. Aspartate aminotransferase (AspAT) is the most extensively studied PLP-containing enzyme. It catalyzes the reversible transamination of dicarboxylic L-amino acids, aspartate and glutamate, and the corresponding 2-oxo acids, oxalacetate and 2-oxoglutarate. Other members of the family included pyruvate aminotransferase, branched-chain amino acid aminotransferase, tyrosine aminotransferase, aromatic aminotransferase, alanine:glyoxylate aminotransferase (AGT), and kynurenine aminotransferase (Vacca, R.A. et al. (1997) J. Biol. Chem. 272:21932-21937).

Primary hyperoxaluria type-1 is an autosomal recessive disorder resulting in a deficiency in the liver-specific peroxisomal enzyme, alanine:glyoxylate aminotransferase-1. The phenotype of the disorder is a deficiency in glyoxylate metabolism. In the absence of AGT, glyoxylate is oxidized to oxalate rather than being transaminated to glycine. The result is the deposition of insoluble calcium

oxalate in the kidneys and urinary tract, ultimately causing renal failure (Lumb, M.J. et al. (1999) J. Biol. Chem. 274:20587-20596).

Kynurenine aminotransferase catalyzes the irreversible transamination of the L-tryptophan metabolite L-kynurenine to form kynurenic acid. The enzyme may also catalyzes the reversible transamination reaction between L-2-aminoadipate and 2-oxoglutarate to produce 2-oxoadipate and L-glutamate. Kynurenic acid is a putative modulator of glutamatergic neurotransmission, thus a deficiency in kynurenine aminotransferase may be associated with pleotrophic effects (Buchli, R. et al. (1995) J. Biol. Chem. 270:29330-29335).

Catechol-*O*-methyltransferase:

Catechol-*O*-methyltransferase (COMT) catalyzes the transfer of the methyl group of *S*-adenosyl-L-methionine (AdoMet; SAM) donor to one of the hydroxyl groups of the catechol substrate (e.g., L-dopa, dopamine, or DBA). Methylation of the 3'-hydroxyl group is favored over methylation of the 4'-hydroxyl group and the membrane bound isoform of COMT is more regiospecific than the soluble form. Translation of the soluble form of the enzyme results from utilization of an internal start codon in a full-length mRNA (1.5 kb) or from the translation of a shorter mRNA (1.3 kb), transcribed from an internal promoter. The proposed S_N2-like methylation reaction requires Mg^{++} and is inhibited by Ca^{++}. The binding of the donor and substrate to COMT occurs sequentially. AdoMet first binds COMT in a Mg^{++}-independent manner, followed by the binding of Mg^{++} and the binding of the catechol substrate.

The amount of COMT in tissues is relatively high compared to the amount of activity normally required, thus inhibition is problematic. Nonetheless, inhibitors have been developed for in vitro use (e.g., gallates, tropolone, U-0521, and 3',4'-dihydroxy-2-methyl-propiophetropolone) and for clinical use (e.g., nitrocatechol-based compounds and tolcapone). Administration of these inhibitors results in the increased half-life of L-dopa and the consequent formation of dopamine. Inhibition of COMT is also likely to increase the half-life of various other catechol-structure compounds, including but not limited to epinephrine/norepinephrine, isoprenaline, rimiterol, dobutamine, fenoldopam, apomorphine, and α-methyldopa. A deficiency in norepinephrine has been linked to clinical depression, hence the use of COMT inhibitors could be usefull in the treatment of depression. COMT inhibitors are generally well tolerated with minimal side effects and are ultimately metabolized in the liver with only minor accumulation of metabolites in the body (Männistö, P.T. and Kaakkola, S. (1999) Pharmacological Reviews 51:593-628).

Copper-zinc superoxide dismutases

Copper-zinc superoxide dismutases are compact homodimeric metalloenzymes involved in cellular defenses against oxidative damage. The enzymes contain one atom of zinc and one atom of copper per subunit and catalyze the dismutation of superoxide anions into O_2 and H_2O_2. The rate of

dismutation is diffusion-limited and consequently enhanced by the presence of favorable electrostatic interactions between the substrate and enzyme active site. Examples of this class of enzyme have been identified in the cytoplasm of all the eukaryotic cells as well as in the periplasm of several bacterial species. Copper-zinc superoxide dismutases are robust enzymes that are highly resistant to proteolytic digestion and denaturing by urea and SDS. In addition to the compact structure of the enzymes, the presence of the metal ions and intrasubunit disulfide bonds is believed to be responsible for enzyme stability. The enzymes undergo reversible denaturation at temperatures as high as 70 °C (Battistoni, A. et al. (1998) J. Biol. Chem. 273:5655-5661).

Overexpression of superoxide dismutase has been implicated in enhancing freezing tolerance of transgenic Alfalfa as well as providing resistance to environmental toxins such as the diphenyl ether herbicide, acifluorfen (McKersie, B.D. et al. (1993) Plant Physiol. 103:1155-1163). In addtion, yeast cells become more resistant to freeze-thaw damage following exposure to hydrogen peroxide which causes the yeast cells to adapt to further peroxide stress by upregulating expression of superoxide dismutases. In this study, mutations to yeast superoxide dismutase genes had a more detrimental effect on freeze-thaw resistance than mutations which affected the regulation of glutathione metabolism, long suspected of being important in determining an organisms survival through the process of cryopreservation (Jong-In Park, J-I. et al. (1998) J. Biol. Chem. 273:22921-22928).

Expression of superoxide dismutase is also associated with Mycobacterium tuberculosis, the organism that causes tuberculosis. Superoxide dismutase is one of the ten major proteins secreted by M. tuberculosis and its expression is upregulated approximately 5-fold in response to oxidative stress. M. tuberculosis expresses almost two orders of magnitude more superoxide dismutase than the nonpathogenic mycobacterium M. smegmatis, and secretes a much higher proportion of the expressed enzyme. The result is the secretion of ~350-fold more enzyme by M. tuberculosis than M. smegmatis, providing substantial resistance to oxidative stress (Harth, G. and Horwitz, M.A. (1999) J. Biol. Chem. 274:4281-4292).

The reduced expression of copper-zinc superoxide dismutases, as well as other enzymes with anti-oxidant capabilities, has been implicated in the early stages of cancer. The expression of copper-zinc superoxide dismutases has been shown to be lower in prostatic intraepithelial neoplasia and prostate carcinomas, compared to normal prostate tissue (Bostwick, D.G. (2000) Cancer 89:123-134).

Phosphodiesterases

Phosphodiesterases make up a class of enzymes which catalyze the hydrolysis of one of the two ester bonds in a phosphodiester compound. Phosphodiesterases are therefore crucial to a variety of cellular processes. Phosphodiesterases include DNA and RNA endonucleases and exonucleases, which are essential for cell growth and replication, and topoisomerases, which break and rejoin

nucleic acid strands during topological rearrangement of DNA. A Tyr-DNA phosphodiesterase functions in DNA repair by hydrolyzing dead-end covalent intermediates formed between topoisomerase I and DNA (Pouliot, J.J. et al. (1999) Science 286:552-555; Yang, S.-W. (1996) Proc. Natl. Acad. Sci. USA 93:11534-11539).

Acid sphingomyelinase is a phosphodiesterase which hydrolyzes the membrane phospholipid sphingomyelin to produce ceramide and phosphorylcholine. Phosphorylcholine is used in the synthesis of phosphatidylcholine, which is involved in numerous intracellular signaling pathways, while ceramide is an essential precursor for the generation of gangliosides, membrane lipids found in high concentration in neural tissue. Defective acid sphingomyelinase leads to a build-up of sphingomyelin molecules in lysosomes, resulting in Niemann-Pick disease (Schuchman, E.H. and S.R. Miranda (1997) Genet. Test. 1:13-19).

Glycerophosphoryl diester phosphodiesterase (also known as glycerophosphodiester phosphodiesterase) is a phosphodiesterase which hydrolyzes deacetylated phospholipid glycerophosphodiesters to produce sn-glycerol-3-phosphate and an alcohol. Glycerophosphocholine, glycerophosphoethanolamine, glycerophosphoglycerol, and glycerophosphoinositol are examples of substrates for glycerophosphoryl diester phosphodiesterases. A glycerophosphoryl diester phosphodiesterase from E. coli has broad specificity for glycerophosphodiester substrates (Larson, T.J. et al. (1983) J. Biol. Chem. 248:5428-5432).

Cyclic nucleotide phosphodiesterases (PDEs) are crucial enzymes in the regulation of the cyclic nucleotides cAMP and cGMP. cAMP and cGMP function as intracellular second messengers to transduce a variety of extracellular signals including hormones, light, and neurotransmitters. PDEs degrade cyclic nucleotides to their corresponding monophosphates, thereby regulating the intracellular concentrations of cyclic nucleotides and their effects on signal transduction. Due to their roles as regulators of signal transduction, PDEs have been extensively studied as chemotherapeutic targets (Perry, M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481; Torphy, J.T. (1998) Am. J. Resp. Crit. Care Med. 157:351-370).

Families of mammalian PDEs have been classified based on their substrate specificity and affinity, sensitivity to cofactors, and sensitivity to inhibitory agents (Beavo, J.A. (1995) Physiol. Rev. 75:725-748; Conti, M. et al. (1995) Endocrine Rev. 16:370-389). Several of these families contain distinct genes, many of which are expressed in different tissues as splice variants. Within PDE families, there are multiple isozymes and multiple splice variants of these isozymes (Conti, M. and S.-L.C. Jin (1999) Prog. Nucleic Acid Res. Mol. Biol. 63:1-38). The existence of multiple PDE families, isozymes, and splice variants is an indication of the variety and complexity of the regulatory pathways involving cyclic nucleotides (Houslay, M.D. and G. Milligan (1997) Trends Biochem. Sci. 22:217-224).

Type 1 PDEs (PDE1s) are Ca^{2+}/calmodulin-dependent and appear to be encoded by at least three different genes, each having at least two different splice variants (Kakkar, R. et al. (1999) Cell Mol. Life Sci. 55:1164-1186). PDE1s have been found in the lung, heart, and brain. Some PDE1 isozymes are regulated in vitro by phosphorylation/dephosphorylation. Phosphorylation of these PDE1 isozymes decreases the affinity of the enzyme for calmodulin, decreases PDE activity, and increases steady state levels of cAMP (Kakkar, supra). PDE1s may provide useful therapeutic targets for disorders of the central nervous system, and the cardiovascular and immune systems due to the involvement of PDE1s in both cyclic nucleotide and calcium signaling (Perry, M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481).

PDE2s are cGMP-stimulated PDEs that have been found in the cerebellum, neocortex, heart, kidney, lung, pulmonary artery, and skeletal muscle (Sadhu, K. et al. (1999) J. Histochem. Cytochem. 47:895-906). PDE2s are thought to mediate the effects of cAMP on catecholamine secretion, participate in the regulation of aldosterone (Beavo, supra), and play a role in olfactory signal transduction (Juilfs, D.M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:3388-3395).

PDE3s have high affinity for both cGMP and cAMP, and so these cyclic nucleotides act as competitive substrates for PDE3s. PDE3s play roles in stimulating myocardial contractility, inhibiting platelet aggregation, relaxing vascular and airway smooth muscle, inhibiting proliferation of T-lymphocytes and cultured vascular smooth muscle cells, and regulating catecholamine-induced release of free fatty acids from adipose tissue. The PDE3 family of phosphodiesterases are sensitive to specific inhibitors such as cilostamide, enoximone, and lixazinone. Isozymes of PDE3 can be regulated by cAMP-dependent protein kinase, or by insulin-dependent kinases (Degerman, E. et al. (1997) J. Biol. Chem. 272:6823-6826).

PDE4s are specific for cAMP; are localized to airway smooth muscle, the vascular endothelium, and all inflammatory cells; and can be activated by cAMP-dependent phosphorylation. Since elevation of cAMP levels can lead to suppression of inflammatory cell activation and to relaxation of bronchial smooth muscle, PDE4s have been studied extensively as possible targets for novel anti-inflammatory agents, with special emphasis placed on the discovery of asthma treatments. PDE4 inhibitors are currently undergoing clinical trials as treatments for asthma, chronic obstructive pulmonary disease, and atopic eczema. All four known isozymes of PDE4 are susceptible to the inhibitor rolipram, a compound which has been shown to improve behavioral memory in mice (Barad, M. et al. (1998) Proc. Natl. Acad. Sci. USA 95:15020-15025). PDE4 inhibitors have also been studied as possible therapeutic agents against acute lung injury, endotoxemia, rheumatoid arthritis, multiple sclerosis, and various neurological and gastrointestinal indications (Doherty, A.M. (1999) Curr. Opin. Chem. Biol. 3:466-473).

PDE5 is highly selective for cGMP as a substrate (Turko, I.V. et al. (1998) Biochemistry

37:4200-4205), and has two allosteric cGMP-specific binding sites (McAllister-Lucas, L.M. et al. (1995) J. Biol. Chem. 270:30671-30679). Binding of cGMP to these allosteric binding sites seems to be important for phosphorylation of PDE5 by cGMP-dependent protein kinase rather than for direct regulation of catalytic activity. High levels of PDE5 are found in vascular smooth muscle, platelets, lung, and kidney. The inhibitor zaprinast is effective against PDE5 and PDE1s. Modification of zaprinast to provide specificity against PDE5 has resulted in sildenafil (VIAGRA; Pfizer, Inc., New York NY), a treatment for male erectile dysfunction (Terrett, N. et al. (1996) Bioorg. Med. Chem. Lett. 6:1819-1824). Inhibitors of PDE5 are currently being studied as agents for cardiovascular therapy (Perry, M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481).

PDE6s, the photoreceptor cyclic nucleotide phosphodiesterases, are crucial components of the phototransduction cascade. In association with the G-protein transducin, PDE6s hydrolyze cGMP to regulate cGMP-gated cation channels in photoreceptor membranes. In addition to the cGMP-binding active site, PDE6s also have two high-affinity cGMP-binding sites which are thought to play a regulatory role in PDE6 function (Artemyev, N.O. et al. (1998) Methods 14:93-104). Defects in PDE6s have been associated with retinal disease. Retinal degeneration in the rd mouse (Yan, W. et al. (1998) Invest. Opthalmol. Vis. Sci. 39:2529-2536), autosomal recessive retinitis pigmentosa in humans (Danciger, M. et al. (1995) Genomics 30:1-7), and rod/cone dysplasia 1 in Irish Setter dogs (Suber, M.L. et al. (1993) Proc. Natl. Acad. Sci. USA 90:3968-3972) have been attributed to mutations in the PDE6B gene.

The PDE7 family of PDEs consists of only one known member having multiple splice variants (Bloom, T.J. and J.A. Beavo (1996) Proc. Natl. Acad. Sci. USA 93:14188-14192). PDE7s are cAMP specific, but little else is known about their physiological function. Although mRNAs encoding PDE7s are found in skeletal muscle, heart, brain, lung, kidney, and pancreas, expression of PDE7 proteins is restricted to specific tissue types (Han, P. et al. (1997) J. Biol. Chem. 272:16152-16157; Perry, M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481). PDE7s are very closely related to the PDE4 family; however, PDE7s are not inhibited by rolipram, a specific inhibitor of PDE4s (Beavo, supra).

PDE8s are cAMP specific, and are closely related to the PDE4 family. PDE8s are expressed in thyroid gland, testis, eye, liver, skeletal muscle, heart, kidney, ovary, and brain. The cAMP-hydrolyzing activity of PDE8s is not inhibited by the PDE inhibitors rolipram, vinpocetine, milrinone, IBMX (3-isobutyl-1-methylxanthine), or zaprinast, but PDE8s are inhibited by dipyridamole (Fisher, D.A. et al. (1998) Biochem. Biophys. Res. Commun. 246:570-577; Hayashi, M. et al. (1998) Biochem. Biophys. Res. Commun. 250:751-756; Soderling, S.H. et al. (1998) Proc. Natl. Acad. Sci. USA 95:8991-8996).

PDE9s are cGMP specific and most closely resemble the PDE8 family of PDEs. PDE9s are

expressed in kidney, liver, lung, brain, spleen, and small intestine. PDE9s are not inhibited by sildenafil (VIAGRA; Pfizer, Inc., New York NY), rolipram, vinpocetine, dipyridamole, or IBMX (3-isobutyl-1-methylxanthine), but they are sensitive to the PDE5 inhibitor zaprinast (Fisher, D.A. et al. (1998) J. Biol. Chem. 273:15559-15564; Soderling, S.H. et al. (1998) J. Biol. Chem. 273:15553-15558).

PDE10s are dual-substrate PDEs, hydrolyzing both cAMP and cGMP. PDE10s are expressed in brain, thyroid, and testis. (Soderling, S.H. et al. (1999) Proc. Natl. Acad. Sci. USA 96:7071-7076; Fujishige, K. et al. (1999) J. Biol. Chem. 274:18438-18445; Loughney, K. et al (1999) Gene 234:109-117).

PDEs are composed of a catalytic domain of about 270-300 amino acids, an N-terminal regulatory domain responsible for binding cofactors, and, in some cases, a hydrophilic C-terminal domain of unknown function (Conti, M. and S.-L.C. Jin (1999) Prog. Nucleic Acid Res. Mol. Biol. 63:1-38). A conserved, putative zinc-binding motif, HDXXHXGXXN, has been identified in the catalytic domain of all PDEs. N-terminal regulatory domains include non-catalytic cGMP-binding domains in PDE2s, PDE5s, and PDE6s; calmodulin-binding domains in PDE1s; and domains containing phosphorylation sites in PDE3s and PDE4s. In PDE5, the N-terminal cGMP-binding domain spans about 380 amino acid residues and comprises tandem repeats of the conserved sequence motif N(R/K)XnFX$_3$DE (McAllister-Lucas, L.M. et al. (1993) J. Biol. Chem. 268:22863-22873). The NKXnD motif has been shown by mutagenesis to be important for cGMP binding (Turko, I.V. et al. (1996) J. Biol. Chem. 271:22240-22244). PDE families display approximately 30% amino acid identity within the catalytic domain; however, isozymes within the same family typically display about 85-95% identity in this region (e.g. PDE4A vs PDE4B). Furthermore, within a family there is extensive similarity (>60%) outside the catalytic domain; while across families, there is little or no sequence similarity outside this domain.

Many of the constituent functions of immune and inflammatory responses are inhibited by agents that increase intracellular levels of cAMP (Verghese, M.W. et al. (1995) Mol. Pharmacol. 47:1164-1171). A variety of diseases have been attributed to increased PDE activity and associated with decreased levels of cyclic nucleotides. For example, a form of diabetes insipidus in mice has been associated with increased PDE4 activity, an increase in low-K_m cAMP PDE activity has been reported in leukocytes of atopic patients, and PDE3 has been associated with cardiac disease.

Many inhibitors of PDEs have been identified and have undergone clinical evaluation (Perry, M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481; Torphy, T.J. (1998) Am. J. Respir. Crit. Care Med. 157:351-370). PDE3 inhibitors are being developed as antithrombotic agents, antihypertensive agents, and as cardiotonic agents useful in the treatment of congestive heart failure. Rolipram, a PDE4 inhibitor, has been used in the treatment of depression, and other inhibitors of

PDE4 are undergoing evaluation as anti-inflammatory agents. Rolipram has also been shown to inhibit lipopolysaccharide (LPS) induced TNF-a which has been shown to enhance HIV-1 replication in vitro. Therefore, rolipram may inhibit HIV-1 replication (Angel, J.B. et al. (1995) AIDS 9:1137-1144). Additionally, rolipram, based on its ability to suppress the production of cytokines such as TNF-a and b and interferon g, has been shown to be effective in the treatment of encephalomyelitis. Rolipram may also be effective in treating tardive dyskinesia and was effective in treating multiple sclerosis in an experimental animal model (Sommer, N. et al. (1995) Nat. Med. 1:244-248; Sasaki, H. et al. (1995) Eur. J. Pharmacol. 282:71-76).

Theophylline is a nonspecific PDE inhibitor used in the treatment of bronchial asthma and other respiratory diseases. Theophylline is believed to act on airway smooth muscle function and in an anti-inflammatory or immunomodulatory capacity in the treatment of respiratory diseases (Banner, K.H. and C.P. Page (1995) Eur. Respir. J. 8:996-1000). Pentoxifylline is another nonspecific PDE inhibitor used in the treatment of intermittent claudication and diabetes-induced peripheral vascular disease. Pentoxifylline is also known to block TNF-a production and may inhibit HIV-1 replication (Angel et al., supra).

PDEs have been reported to affect cellular proliferation of a variety of cell types (Conti et al. (1995) Endocrine Rev. 16:370-389) and have been implicated in various cancers. Growth of prostate carcinoma cell lines DU145 and LNCaP was inhibited by delivery of cAMP derivatives and PDE inhibitors (Bang, Y.J. et al. (1994) Proc. Natl. Acad. Sci. USA 91:5330-5334). These cells also showed a permanent conversion in phenotype from epithelial to neuronal morphology. It has also been suggested that PDE inhibitors have the potential to regulate mesangial cell proliferation (Matousovic, K. et al. (1995) J. Clin. Invest. 96:401-410) and lymphocyte proliferation (Joulain, C. et al. (1995) J. Lipid Mediat. Cell Signal. 11:63-79). A cancer treatment has been described that involves intracellular delivery of PDEs to particular cellular compartments of tumors, resulting in cell death (Deonarain, M.P. and A.A. Epenetos (1994) Br. J. Cancer 70:786-794).

Phosphotriesterases

Phosphotriesterases (PTE, paraoxonases) are enzymes that hydrolyze toxic organophosphorus compounds and have been isolated from a variety of tissues. The enzymes appear to be lacking in birds and insects and abundant in mammals, explaining the reduced tolerance of birds and insects to organophosphorus compound (Vilanova, E. and Sogorb, M.A. (1999) Crit. Rev. Toxicol. 29:21-57). Phosphotriesterases play a central role in the detoxification of insecticides by mammals. Phosphotriesterase activity varies among individuals and is lower in infants than adults. Knockout mice are markedly more sensitive to the organophosphate-based toxins diazoxon and chlorpyrifos oxon (Furlong, C.E., et al. (2000) Neurotoxicology 21:91-100). PTEs have attracted interest as enzymes capable of the detoxification of organophosphate-containing chemical waste and warfare

reagents (e.g., parathion), in addition to pesticides and insecticides. Some studies have also implicated phosphotriesterase in atherosclerosis and diseases involving lipoprotein metabolism.

Thioesterases

Two soluble thioesterases involved in fatty acid biosynthesis have been isolated from mammalian tissues, one which is active only toward long-chain fatty-acyl thioesters and one which is active toward thioesters with a wide range of fatty-acyl chain-lengths. These thioesterases catalyze the chain-terminating step in the *de novo* biosynthesis of fatty acids. Chain termination involves the hydrolysis of the thioester bond which links the fatty acyl chain to the 4'-phosphopantetheine prosthetic group of the acyl carrier protein (ACP) subunit of the fatty acid synthase (Smith, S. (1981a) Methods Enzymol. 71:181-188; Smith, S. (1981b) Methods Enzymol. 71:188-200).

E. coli contains two soluble thioesterases, thioesterase I which is active only toward long-chain acyl thioesters, and thioesterase II (TEII) which has a broad chain-length specificity (Naggert, J. et al. (1991) J. Biol. Chem. 266:11044-11050). E. coli TEII does not exhibit sequence similarity with either of the two types of mammalian thioesterases which function as chain-terminating enzymes in *de novo* fatty acid biosynthesis. Unlike the mammalian thioesterases, E. coli TEII lacks the characteristic serine active site gly-X-ser-X-gly sequence motif and is not inactivated by the serine modifying agent diisopropyl fluorophosphate. However, modification of histidine 58 by iodoacetamide and diethylpyrocarbonate abolished TEII activity. Overexpression of TEII did not alter fatty acid content in E. coli, which suggests that it does not function as a chain-terminating enzyme in fatty acid biosynthesis (Naggert et al., supra). For that reason, Naggert et al. (supra) proposed that the physiological substrates for E. coli TEII may be coenzyme A (CoA)-fatty acid esters instead of ACP-phosphopanthetheine-fatty acid esters.

Carboxylesterases

Mammalian carboxylesterases constitute a multigene family expressed in a variety of tissues and cell types. Isozymes have significant sequence homology and are classified primarily on the basis of amino acid sequence. Acetylcholinesterase, butyrylcholinesterase, and carboxylesterase are grouped into the serine super family of esterases (B-esterases). Other carboxylesterases included thyroglobulin, thrombin, Factor IX, gliotactin, and plasminogen. Carboxylesterases catalyze the hydrolysis of ester- and amide- groups from molecules and are involved in detoxification of drugs, environmental toxins, and carcinogens. Substrates for carboxylesterases include short- and long-chain acyl-glycerols, acylcarnitine, carbonates, dipivefrin hydrochloride, cocaine, salicylates, capsaicin, palmitoyl-coenzyme A, imidapril, haloperidol, pyrrolizidine alkaloids, steroids, p-nitrophenyl acetate, malathion, butanilicaine, and isocarboxazide. The enzymes often demonstrate low substrate specificity. Carboxylesterases are also important for the conversion of prodrugs to their respective free acids, which may be the active form of the drug (e.g., lovastatin, used to lower blood

cholesterol) (reviewed in Satoh, T. and Hosokawa, M. (1998) Annu. Rev. Pharmacol. Toxicol.38:257-288).

Neuroligins are a class of molecules that (i) have N-terminal signal sequences, (ii) resemble cell-surface receptors, (iii) contain carboxylesterase domains, (iv) are highly expressed in the brain, and (v) bind to neurexins in a calcium-dependent manner. Despite the homology to carboxylesterases, neuroligins lack the active site serine residue, implying a role in substrate binding rather than catalysis (Ichtchenko, K. et al. (1996) J. Biol. Chem. 271:2676-2682).

Squalene epoxidase

Squalene epoxidase (squalene monooxygenase, SE) is a microsomal membrane-bound, FAD-dependent oxidoreductase that catalyzes the first oxygenation step in the sterol biosynthetic pathway of eukaryotic cells. Cholesterol is an essential structural component of cytoplasmic membranes acquired via the LDL receptor-mediated pathway or the biosynthetic pathway. In the latter case, all 27 carbon atoms in the cholesterol molecule are derived from acetyl-CoA (Stryer, L., supra). SE converts squalene to 2,3(*S*)-oxidosqualene, which is then converted to lanosterol and then cholesterol. The steps involved in cholesterol biosynthesis are summarized below (Stryer, L (1988) Biochemistry. W.H Freeman and Co., Inc. New York. pp. 554-560 and Sakakibara, J. et al. (1995) 270:17-20):

acetate (from Acetyl-CoA) → 3-hydoxy-3-methyl-glutaryl CoA → mevalonate → 5-phosphomevalonate → 5-pyrophosphomevalonate → isopentenyl pyrophosphate → dimethylallyl pyrophosphate → geranyl pyrophosphate → farnesyl pyrophosphate → squalene → squalene epoxide → lanosterol → cholesterol

While cholesterol is essential for the viability of eukaryotic cells, inordinately high serum cholesterol levels results in the formation of atherosclerotic plaques in the arteries of higher organisms. This deposition of highly insoluble lipid material onto the walls of essential blood vessels (e.g., coronary arteries) results in decreased blood flow and potential necrosis of the tissues deprived of adequate blood flow. HMG-CoA reductase is responsible for the conversion of 3-hydroxyl-3-methyl-glutaryl CoA (HMG-CoA) to mevalonate, which represents the first committed step in cholesterol biosynthesis. HMG-CoA is the target of a number of pharmaceutical compounds designed to lower plasma cholesterol levels. However, inhibition of MHG-CoA also results in the reduced synthesis of non-sterol intermediates (e.g., mevalonate) required for other biochemical pathways. SE catalyzes a rate-limiting reaction that occurs later in the sterol synthesis pathway and cholesterol in the only end product of the pathway following the step catalyzed by SE. As a result, SE is the ideal target for the design of anti-hyperlipidemic drugs that do not cause a reduction in other necessary intermediates (Nakamura, Y. et al. (1996) 271:8053-8056).

Epoxide hydrolases

Epoxide hydrolases catalyze the addition of water to epoxide-containing compounds, thereby hydrolyzing epoxides to their corresponding 1,2-diols. They are related to bacterial haloalkane dehalogenases and show sequence similarity to other members of the α/β hydrolase fold family of enzymes (e.g., bromoperoxidase A2 from Streptomyces aureofaciens, hydroxymuconic semialdehyde hydrolases from Pseudomonas putida, and haloalkane dehalogenase from Xanthobacter autotrophicus). Epoxide hydrolases are ubiquitous in nature and have been found in mammals, invertebrates, plants, fungi, and bacteria. This family of enzymes is important for the detoxification of xenobiotic epoxide compounds which are often highly electrophilic and destructive when introduced into an organism. Examples of epoxide hydrolase reactions include the hydrolysis of cis-9,10-epoxyoctadec-9(Z)-enoic acid (leukotoxin) to form its corresponding diol, threo-9,10-dihydroxyoctadec-12(Z)-enoic acid (leukotoxin diol), and the hydrolysis of cis-12,13-epoxyoctadec-9(Z)-enoic acid (isoleukotoxin) to form its corresponding diol threo-12,13-dihydroxyoctadec-9(Z)-enoic acid (isoleukotoxin diol). Leukotoxins alter membrane permeability and ion transport and cause inflammatory responses. In addition, epoxide carcinogens are known to be produced by cytochrome P450 as intermediates in the detoxification of drugs and environmental toxins.

The enzymes possess a catalytic triad composed of Asp (the nucleophile), Asp (the histidine-supporting acid), and His (the water-activating histidine). The reaction mechanism of epoxide hydrolase proceeds via a covalently bound ester intermediate initiated by the nucleophilic attack of one of the Asp residues on the primary carbon atom of the epoxide ring of the target molecule, leading to a covalently bound ester intermediate (Michael Arand, M. et al. (1996) J. Biol. Chem. 271:4223-4229; Rink, R. et al. (1997) J. Biol. Chem. 272:14650-14657; Argiriadi, M.A. et al. (2000) J. Biol. Chem. 275:15265-15270).

Enzymes involved in tyrosine catalysis

The degradation of the amino acid tyrosine to either succinate and pyruvate or fumarate and acetoacetate, requires a large number of enzymes and generates a large number of intermediate compounds. In addition, many xenobiotic compounds may be metabolized using one or more reactions that are part of the tyrosine catabolic pathway. While the pathway has been studied primarily in bacteria, tyrosine degradation is known to occur in a variety of organisms and is likely to involve many of the same biological reactions.

The enzymes involved in the degradation of tyrosine to succinate and pyruvate (e.g., in *Arthrobacter* species) include 4-hydroxyphenylpyruvate oxidase, 4-hydroxyphenylacetate 3-hydroxylase, 3,4-dihydroxyphenylacetate 2,3-dioxygenase, 5-carboxymethyl-2-hydroxymuconic semialdehyde dehydrogenase, *trans,cis*-5-carboxymethyl-2-hydroxymuconate isomerase, homoprotocatechuate isomerase/decarboxylase, *cis*-2-oxohept-3-ene-1,7-dioate hydratase,

2,4-dihydroxyhept-*trans*-2-ene-1,7-dioate aldolase, and succinic semialdehyde dehydrogenase.

The enzymes involved in the degradation of tyrosine to fumarate and acetoacetate (e.g., in *Pseudomonas* species) include 4-hydroxyphenylpyruvate dioxygenase, homogentisate 1,2-dioxygenase, maleylacetoacetate isomerase, and fumarylacetoacetase. 4-hydroxyphenylacetate 1-hydroxylase may also be involved if intermediates from the succinate/pyruvate pathway are accepted.

Additional enzymes associated with tyrosine metabolism in different organisms include 4-chlorophenylacetate-3,4-dioxygenase, aromatic aminotransferase, 5-oxopent-3-ene-1,2,5-tricarboxylate decarboxylase, 2-oxo-hept-3-ene-1,7-dioate hydratase, and 5-carboxymethyl-2-hydroxymuconate isomerase (Ellis, L.B.M. et al. (1999) Nucleic Acids Res. 27:373-376; Wackett, L.P. and Ellis, L.B.M. (1996) J. Microbiol. Meth. 25:91-93; and Schmidt, M. (1996) Amer. Soc. Microbiol. News 62:102).

In humans, acquired or inherited genetic defects in enzymes of the tyrosine degradation pathway may result in hereditary tyrosinemia. One form of this disease, hereditary tyrosinemia 1 (HT1) is caused by a deficiency in the enzyme fumarylacetoacetate hydrolase, the last enzyme in the pathway in organisms that metabolize tyrosine to fumarate and acetoacetate. HT1 is characterized by progressive liver damage beginning at infancy, and increased risk for liver cancer (Endo, F. et al. (1997) J. Biol. Chem. 272:24426-24432).

The discovery of new drug metabolizing enzymes and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of autoimmune/inflammatory, cell proliferative, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of drug metabolizing enzymes.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, drug metabolizing enzymes, referred to collectively as "DME" and individually as "DME-1," "DME-2," "DME-3," "DME-4," "DME-5," "DME-6," "DME-7," "DME-8," "DME-9," "DME-10," "DME-11," "DME-12," "DME-13," "DME-14," "DME-15," "DME-16," "DME-17," "DME-18," "DME-19," "DME-20," "DME-21," "DME-22," "DME-23," and "DME-24." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting

of SEQ ID NO:1-24, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-24.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-24. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:25-48.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino

acid sequence selected from the group consisting of SEQ ID NO:1-24, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:25-48, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:25- 48, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:25-48, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:25-48, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:25-48, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:25-48, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-24. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional DME, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional DME, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or

condition associated with overexpression of functional DME, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:25-48, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID

NO:25-48, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:25-48, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:25-48, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:25-48, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for each polypeptide of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of each polypeptide sequence, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of each polypeptide.

Table 4 lists the cDNA and genomic DNA fragments which were used to assemble each polynucleotide sequence, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for each polynucleotide of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"DME" refers to the amino acid sequences of substantially purified DME obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of DME. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of DME either by directly interacting with DME or by acting on components of the biological pathway in which DME participates.

An "allelic variant" is an alternative form of the gene encoding DME. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding DME include those sequences with deletions,

insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as DME or a polypeptide with at least one functional characteristic of DME. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding DME, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding DME. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent DME. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of DME is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of DME. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of DME either by directly interacting with DME or by acting on components of the biological pathway in which DME participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind DME polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired.

Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic DME, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding DME or fragments of DME may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium

dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide.

Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

A "fragment" is a unique portion of DME or the polynucleotide encoding DME which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:25-48 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:25-48, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:25-48 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:25-48 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:25-48 and the region of SEQ ID NO:25-48 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-24 is encoded by a fragment of SEQ ID NO:25-48. A fragment of SEQ ID NO:1-24 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-24. For example, a fragment of SEQ ID NO:1-24 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-24. The precise length of a fragment of SEQ ID NO:1-24 and the region of SEQ ID NO:1-24 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A

"full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press,

Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of DME which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of DME which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of DME. For example, modulation

may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of DME.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an DME may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of DME.

"Probe" refers to nucleic acid sequences encoding DME, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have

been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing DME, nucleic acids encoding DME, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides

by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least

98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human drug metabolizing enzymes (DME), the polynucleotides encoding DME, and the use of these compositions for the diagnosis, treatment, or prevention of autoimmune/inflammatory, cell proliferative, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte

polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog.

Table 3 shows various structural features of each of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of each polypeptide of the invention, and these properties establish that the claimed polypeptides are drug metabolizing enzymes. The algorithms and parameters for the analysis of SEQ ID NO:1-24 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:25-48 or that distinguish between SEQ ID NO:25-48 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5′) and stop (3′) positions of the cDNA and genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 6537030H1 is the identification number of an Incyte cDNA sequence, and (OVARDIN02) is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 70614021V1). Alternatively, the identification numbers in column 5

may refer to GenBank cDNAs or ESTs (e.g., g758933) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, g5091644.v113.gs_1.1.nt.edit is the identification number of a Genscan-predicted coding sequence, with g5091644 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses DME variants. A preferred DME variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the DME amino acid sequence, and which contains at least one functional or structural characteristic of DME.

The invention also encompasses polynucleotides which encode DME. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:25-48, which encodes DME. The polynucleotide sequences of SEQ ID NO:25-48, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding DME. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding DME. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:25-48 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting

of SEQ ID NO:25-48. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of DME.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding DME, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring DME, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode DME and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring DME under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding DME or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding DME and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode DME and DME derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding DME or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:25-48 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or

combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding DME may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of

sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode DME may be cloned in recombinant DNA molecules that direct expression of DME, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express DME.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter DME-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of DME, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are

optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding DME may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, DME itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp.55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of DME, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active DME, the nucleotide sequences encoding DME or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding DME. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding DME. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding DME and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding DME and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding DME. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al, (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding DME. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding DME can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding DME into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of DME are needed, e.g. for the production of

antibodies, vectors which direct high level expression of DME may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of DME. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of DME. Transcription of sequences encoding DME may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding DME may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses DME in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of DME in cell lines is preferred. For example, sequences encoding DME can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the

introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding DME is inserted within a marker gene sequence, transformed cells containing sequences encoding DME can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding DME under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding DME and that express DME may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of DME using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques

include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on DME is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding DME include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding DME, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding DME may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode DME may be designed to contain signal sequences which direct secretion of DME through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct

modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding DME may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric DME protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of DME activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the DME encoding sequence and the heterologous protein sequence, so that DME may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled DME may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

DME of the present invention or fragments thereof may be used to screen for compounds that specifically bind to DME. At least one and up to a plurality of test compounds may be screened for specific binding to DME. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of DME, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which DME binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express DME, either as a secreted

protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing DME or cell membrane fractions which contain DME are then contacted with a test compound and binding, stimulation, or inhibition of activity of either DME or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with DME, either in solution or affixed to a solid support, and detecting the binding of DME to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

DME of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of DME. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for DME activity, wherein DME is combined with at least one test compound, and the activity of DME in the presence of a test compound is compared with the activity of DME in the absence of the test compound. A change in the activity of DME in the presence of the test compound is indicative of a compound that modulates the activity of DME. Alternatively, a test compound is combined with an in vitro or cell-free system comprising DME under conditions suitable for DME activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of DME may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding DME or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell

blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding DME may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding DME can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding DME is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress DME, e.g., by secreting DME in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of DME and drug metabolizing enzymes. In addition, the expression of DME is closely associated with brain, breast, prostate, ovary, testicle, bone, blood, kidney, lung, thyroid, and gastrointestinal tissues; Crohn's disease; breast, sigmoid mesentery, and ureter tumors; and cancers of the lung, prostate, bone, and blood. Therefore, DME appears to play a role in autoimmune/inflammatory, cell proliferative, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders. In the treatment of disorders associated with increased DME expression or activity, it is desirable to decrease the expression or activity of DME. In the treatment of disorders associated with decreased DME expression or activity, it is desirable to increase the expression or activity of DME.

Therefore, in one embodiment, DME or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of DME. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED),

bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma; melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an endocrine disorder, such as disorders of the hypothalamus and pituitary resulting from lesions such as primary brain tumors, adenomas, infarction associated with pregnancy, hypophysectomy, aneurysms, vascular malformations, thrombosis, infections, immunological disorders, and complications due to head trauma; disorders associated with hypopituitarism including hypogonadism, Sheehan syndrome, diabetes insipidus, Kallman's disease, Hand-Schuller-Christian disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and dwarfism; disorders associated with hyperpituitarism including acromegaly, giantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma; disorders associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection, subacute thyroiditis associated with viral infection, autoimmune thyroiditis (Hashimoto's disease), and cretinism; disorders associated with hyperthyroidism including thyrotoxicosis and its various forms, Grave's disease, pretibial myxedema, toxic multinodular goiter,

thyroid carcinoma, and Plummer's disease; disorders associated with hyperparathyroidism including Conn disease (chronic hypercalemia); pancreatic disorders such as Type I or Type II diabetes mellitus and associated complications; disorders associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis, amyloidosis, hypokalemia, Cushing's disease, Liddle's syndrome, and Arnold-Healy-Gordon syndrome, pheochromocytoma tumors, and Addison's disease; disorders associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, endometriosis, perturbations of the menstrual cycle, polycystic ovarian disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, hypergonadal disorders associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5 α-reductase, and gynecomastia; an eye disorder, such as conjunctivitis, keratoconjunctivitis sicca, keratitis, episcleritis, iritis, posterior uveitis, glaucoma, amaurosis fugax, ischemic optic neuropathy, optic neuritis, Leber's hereditary optic neuropathy, toxic optic neuropathy, vitreous detachment, retinal detachment, cataract, macular degeneration, central serous chorioretinopathy, retinitis pigmentosa, melanoma of the choroid, retrobulbar tumor, and chiasmal tumor; a metabolic disorder, such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, diabetes, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, lipid myopathies, lipodystrophies, lysosomal storage diseases, Menkes syndrome, occipital horn syndrome, mannosidosis, neuraminidase deficiency, obesity, pentosuria phenylketonuria, pseudovitamin D-deficiency rickets; hypocalcemia, hypophosphatemia, and postpubescent cerebellar ataxia, tyrosinemia, and a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, hereditary hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic

encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha$_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas.

In another embodiment, a vector capable of expressing DME or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of DME including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified DME in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of DME including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of DME may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of DME including, but not limited to, those listed above.

In a further embodiment, an antagonist of DME may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of DME. Examples of such disorders include, but are not limited to, those autoimmune/inflammatory, cell proliferative, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders, described above. In one aspect, an antibody which specifically binds DME may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express DME.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding DME may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of DME including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of DME may be produced using methods which are generally known in the art. In particular, purified DME may be used to produce antibodies or to screen libraries of

pharmaceutical agents to identify those which specifically bind DME. Antibodies to DME may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with DME or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to DME have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of DME amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to DME may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce DME-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte

population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for DME may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between DME and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering DME epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for DME. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of DME-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple DME epitopes, represents the average affinity, or avidity, of the antibodies for DME. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular DME epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the DME-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of DME, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation

of DME-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding DME, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding DME. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding DME. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Cli. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding DME may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular

parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in DME expression or regulation causes disease, the expression of DME from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in DME are treated by constructing mammalian expression vectors encoding DME and introducing these vectors by mechanical means into DME-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of DME include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). DME may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding DME from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to DME expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding DME under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding DME to cells which have one or more genetic abnormalities with respect to the expression of DME. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding DME to target cells which have one or more genetic abnormalities with respect to the expression of DME. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing DME to cells of the central nervous system, for which HSV has a

tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding DME to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for DME into the alphavirus genome in place of the capsid-coding region results in the production of a large number of DME-coding RNAs and the synthesis of high levels of DME in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of DME into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions

-10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding DME.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding DME. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous

endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding DME. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased DME expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding DME may be therapeutically useful, and in the treament of disorders associated with decreased DME expression or activity, a compound which specifically promotes expression of the polynucleotide encoding DME may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding DME is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding DME are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding DME. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides,

ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of DME, antibodies to DME, and mimetics, agonists, antagonists, or inhibitors of DME.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of

macromolecules comprising DME or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, DME or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example DME or fragments thereof, antibodies of DME, and agonists, antagonists or inhibitors of DME, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells,

conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind DME may be used for the diagnosis of disorders characterized by expression of DME, or in assays to monitor patients being treated with DME or agonists, antagonists, or inhibitors of DME. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for DME include methods which utilize the antibody and a label to detect DME in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring DME, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of DME expression. Normal or standard values for DME expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to DME under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of DME expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding DME may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of DME may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of DME, and to monitor regulation of DME levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding DME or closely related molecules may be used to identify nucleic acid sequences which encode DME. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding DME, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the DME encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:25-48 or from

genomic sequences including promoters, enhancers, and introns of the DME gene.

Means for producing specific hybridization probes for DNAs encoding DME include the cloning of polynucleotide sequences encoding DME or DME derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding DME may be used for the diagnosis of disorders associated with expression of DME. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and

neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an endocrine disorder, such as disorders of the hypothalamus and pituitary resulting from lesions such as primary brain tumors, adenomas, infarction associated with pregnancy, hypophysectomy, aneurysms, vascular malformations, thrombosis, infections, immunological disorders, and complications due to head trauma; disorders associated with hypopituitarism including hypogonadism, Sheehan syndrome, diabetes insipidus, Kallman's disease, Hand-Schuller-Christian disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and dwarfism; disorders associated with hyperpituitarism including acromegaly, giantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma; disorders associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection, subacute thyroiditis associated with viral infection, autoimmune thyroiditis (Hashimoto's disease), and cretinism; disorders associated with hyperthyroidism including thyrotoxicosis and its various forms, Grave's disease, pretibial myxedema, toxic multinodular goiter, thyroid carcinoma, and Plummer's disease; disorders associated with hyperparathyroidism including Conn disease (chronic hypercalemia); pancreatic disorders such as Type I or Type II diabetes mellitus and associated complications; disorders associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis, amyloidosis, hypokalemia, Cushing's disease, Liddle's syndrome, and Arnold-Healy-Gordon syndrome, pheochromocytoma tumors, and Addison's disease; disorders associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, endometriosis, perturbations of the menstrual cycle, polycystic ovarian disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, hypergonadal disorders associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5 α-reductase, and gynecomastia; an eye disorder, such as conjunctivitis, keratoconjunctivitis sicca, keratitis, episcleritis, iritis, posterior uveitis, glaucoma, amaurosis fugax, ischemic optic neuropathy, optic neuritis, Leber's hereditary optic neuropathy, toxic optic neuropathy, vitreous detachment, retinal detachment, cataract, macular degeneration, central serous chorioretinopathy, retinitis pigmentosa, melanoma of the choroid, retrobulbar tumor, and chiasmal tumor; a metabolic disorder, such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, diabetes, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism,

hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, lipid myopathies, lipodystrophies, lysosomal storage diseases, Menkes syndrome, occipital horn syndrome, mannosidosis, neuraminidase deficiency, obesity, pentosuria phenylketonuria, pseudovitamin D-deficiency rickets; hypocalcemia, hypophosphatemia, and postpubescent cerebellar ataxia, tyrosinemia, and a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, hereditary hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, $alpha_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas. The polynucleotide sequences encoding DME may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered DME expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding DME may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding DME may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding DME in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of DME, a normal or standard profile for expression is established. This may be accomplished by

combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding DME, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding DME may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding DME, or a fragment of a polynucleotide complementary to the polynucleotide encoding DME, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding DME may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding DME are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the

oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of DME include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, DME, fragments of DME, or antibodies specific for DME may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions

and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared

with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for DME to quantify the levels of DME expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be

useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding DME may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat.

Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding DME on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, DME, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between DME and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with DME, or fragments thereof, and washed. Bound DME is then detected by methods well known in the art. Purified DME can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding DME specifically compete with a test compound for binding DME. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with DME.

In additional embodiments, the nucleotide sequences which encode DME may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. Nos. 60/176,139, 60/177,443, and 60/178,574, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra,

units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the

ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate

references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:25-48. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative drug metabolizing enzymes were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode drug metabolizing enzymes, the encoded polypeptides were analyzed by querying against PFAM models for drug metabolizing enzymes. Potential drug metabolizing enzymes were also identified by homology to Incyte cDNA sequences that had been annotated as drug metabolizing enzymes. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example

III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of DME Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:25-48 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched

SEQ ID NO:25-48 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, or human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair

(HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding DME are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding DME. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of DME Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing

primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:25-48 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope

slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522 , incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores.

Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the DME-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring DME. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of DME. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the DME-encoding transcript.

XII. Expression of DME

Expression and purification of DME is achieved using bacterial or virus-based expression systems. For expression of DME in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express DME upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of DME in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding DME by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, DME is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from DME at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified DME obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, where applicable.

XIII. Functional Assays

DME function is assessed by expressing the sequences encoding DME at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected

into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of DME on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding DME and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding DME and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of DME Specific Antibodies

DME substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the DME amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-

Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-DME activity by, for example, binding the peptide or DME to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring DME Using Specific Antibodies

Naturally occurring or recombinant DME is substantially purified by immunoaffinity chromatography using antibodies specific for DME. An immunoaffinity column is constructed by covalently coupling anti-DME antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing DME are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of DME (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/DME binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and DME is collected.

XVI. Identification of Molecules Which Interact with DME

DME, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled DME, washed, and any wells with labeled DME complex are assayed. Data obtained using different concentrations of DME are used to calculate values for the number, affinity, and association of DME with the candidate molecules.

Alternatively, molecules interacting with DME are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

DME may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Demonstration of DME Activity

Cytochrome P450 activity of DME is measured using the 4-hydroxylation of aniline. Aniline is converted to 4-aminophenol by the enzyme, and has an absorption maximum at 630 nm (Gibson and Skett, supra). This assay is a convenient measure, but underestimates the total hydroxylation,

which also occurs at the 2- and 3- positions. Assays are performed at 37 °C and contain an aliquot of the enzyme and a suitable amount of aniline (approximately 2 mM) in reaction buffer. For this reaction, the buffer must contain NADPH or an NADPH-generating cofactor system. One formulation for this reaction buffer includes 85 mM Tris pH 7.4, 15 mM $MgCl_2$, 50 mM nicotinamide, 40 mg trisodium isocitrate, and 2 units isocitrate dehydrogenase, with 8 mg $NADP^+$ added to a 10 mL reaction buffer stock just prior to assay. Reactions are carried out in an optical cuvette, and the absorbance at 630 nm is measured. The rate of increase in absorbance is proportional to the enzyme activity in the assay. A standard curve can be constructed using known concentrations of 4-aminophenol.

1α,25-dihydroxyvitamin D 24-hydroxylase activity of DME is determined by monitoring the conversion of ^{3}H-labeled 1α,25-dihydroxyvitamin D (1α,25$(OH)_2$D) to 24,25-dihydroxyvitamin D (24,25$(OH)_2$D) in transgenic rats expressing DME. 1 μg of 1α,25$(OH)_2$D dissolved in ethanol (or ethanol alone as a control) is administered intravenously to approximately 6-week-old male transgenic rats expressing DME or otherwise identical control rats expressing either a defective variant of DME or not expressing DME. The rats are killed by decapitation after 8 hrs, and the kidneys are rapidly removed, rinsed, and homogenized in 9 volumes of ice-cold buffer (15 mM Tris-acetate (pH 7.4), 0.19 M sucrose, 2 mM magnesium acetate, and 5 mM sodium succinate). A portion (e.g., 3 ml) of each homogenate is then incubated with 0.25 nM 1α,25$(OH)_2$[1-^{3}H]D, with a specific activity of approximately 3.5 GBq/mmol, for 15 min at 37 °C under oxygen with constant shaking. Total lipids are extracted as described (Bligh, E.G. and Dyer, W.J. (1959) Can. J. Biochem. Physiol. 37: 911-917) and the chloroform phase is analyzed by HPLC using a FINEPAK SIL column (JASCO, Tokyo, Japan) with a *n*-hexane/chloroform/methanol (10:2.5:1.5) solvent system at a flow rate of 1 ml/min. In the alternative, the chloroform phase is analyzed by reverse phase HPLC using a J SPHERE ODS-AM column (YMC Co. Ltd., Kyoto, Japan) with an acetonitrile buffer system (40 to 100%, in water, in 30 min) at a flow rate of 1 ml/min. The eluates are collected in fractions of 30 seconds (or less) and the amount of ^{3}H present in each fraction is measured using a scintillation counter. By comparing the chromatograms of control samples (i.e., samples comprising 1α,25-dihydroxyvitamin D or 24,25-dihydroxyvitamin D (24,25$(OH)_2$D), with the chromatograms of the reaction products, the relative mobilities of the substrate (1α,25$(OH)_2$[1-^{3}H]D) and product (24,25$(OH)_2$[1-^{3}H]D) are determined and correlated with the fractions collected. The amount of 24,25$(OH)_2$[1-^{3}H]D produced in control rats is subtracted from that of transgenic rats expressing DME. The difference in the production of 24,25$(OH)_2$[1-^{3}H]D in the transgenic and control animals is proportional to the amount of 25-hydrolase activity of DME present in the sample. Confirmation of the identity of the substrate and product(s) is confirmed by means of mass spectroscopy (Miyamoto, Y. et al. (1997) J. Biol. Chem. 272:14115-14119).

Flavin-containing monooxygenase activity of DME is measured by chromatographic analysis of metabolic products. For example, Ring, B. J. et al. (1999; Drug Metab. Dis. 27:1099-1103) incubated FMO in 0.1 M sodium phosphate buffer (pH 7.4 or 8.3) and 1 mM NADPH at 37°C, stopped the reaction with an organic solvent, and determined product formation by HPLC. Alternatively, activity is measured by monitoring oxygen uptake using a Clark-type electrode. For example, Ziegler, D. M. and Poulsen, L. L. (1978; Methods Enzymol. 52:142-151) incubated the enzyme at 37°C in an NADPH-generating cofactor system (similar to the one described above) containing the substrate methimazole. The rate of oxygen uptake is proportional to enzyme activity.

UDP glucuronyltransferase activity of DME is measured using a colorimetric determination of free amine groups (Gibson and Skett, supra). An amine-containing substrate, such as 2-aminophenol, is incubated at 37°C with an aliquot of the enzyme in a reaction buffer containing the necessary cofactors (40 mM Tris pH 8.0, 7.5 mM $MgCl_2$, 0.025% Triton X-100, 1 mM ascorbic acid, 0.75 mM UDP-glucuronic acid). After sufficient time, the reaction is stopped by addition of ice-cold 20% trichloroacetic acid in 0.1 M phosphate buffer pH 2.7, incubated on ice, and centrifuged to clarify the supernatant. Any unreacted 2-aminophenol is destroyed in this step. Sufficient freshly-prepared sodium nitrite is then added; this step allows formation of the diazonium salt of the glucuronidated product. Excess nitrite is removed by addition of sufficient ammonium sulfamate, and the diazonium salt is reacted with an aromatic amine (for example, N-naphthylethylene diamine) to produce a colored azo compound which can be assayed spectrophotometrically (at 540 nm for the example). A standard curve can be constructed using known concentrations of aniline, which will form a chromophore with similar properties to 2-aminophenol glucuronide.

Glutathione S-transferase activity of DME is measured using a model substrate, such as 2,4-dinitro-1-chlorobenzene, which reacts with glutathione to form a product, 2,4-dinitrophenyl-glutathione, that has an absorbance maximum at 340 nm. It is important to note that GSTs have differing substrate specificities, and the model substrate should be selected based on the substrate preferences of the GST of interest. Assays are performed at ambient temperature and contain an aliquot of the enzyme in a suitable reaction buffer (for example, 1 mM glutathione, 1 mM dinitrochlorobenzene, 90 mM potassium phosphate buffer pH 6.5). Reactions are carried out in an optical cuvette, and the absorbance at 340 nm is measured. The rate of increase in absorbance is proportional to the enzyme activity in the assay.

N-acyltransferase activity of DME is measured using radiolabeled amino acid substrates and measuring radiolabel incorporation into conjugated products. Enzyme is incubated in a reaction buffer containing an unlabeled acyl-CoA compound and radiolabeled amino acid, and the radiolabeled acyl-conjugates are separated from the unreacted amino acid by extraction into n-butanol or other appropriate organic solvent. For example, Johnson, M. R. et al. (1990; J. Biol.

Chem. 266:10227-10233) measured bile acid-CoA:amino acid N-acyltransferase activity by incubating the enzyme with cholyl-CoA and ^{3}H-glycine or ^{3}H-taurine, separating the tritiated cholate conjugate by extraction into n-butanol, and measuring the radioactivity in the extracted product by scintillation. Alternatively, N-acyltransferase activity is measured using the spectrophotometric determination of reduced CoA (CoASH) described below.

N-acetyltransferase activity of DME is measured using the transfer of radiolabel from [^{14}C]acetyl-CoA to a substrate molecule (for example, see Deguchi, T. (1975) J. Neurochem. 24:1083-5). Alternatively, a spectrophotometric assay based on DTNB (5,5'-dithio-bis(2-nitrobenzoic acid; Ellman's reagent) reaction with CoASH may be used. Free thiol-containing CoASH is formed during N-acetyltransferase catalyzed transfer of an acetyl group to a substrate. CoASH is detected using the absorbance of DTNB conjugate at 412 nm (De Angelis, J. et al. (1997) J. Biol. Chem. 273:3045-3050). Enzyme activity is proportional to the rate of radioactivity incorporation into substrate, or the rate of absorbance increase in the spectrophotometric assay.

Catechol-*O*-methyltransferase activity of DME is measured in a reaction mixture consisting of 50 mM Tris-HCl (pH 7.4), 1.2 mM $MgCl_2$, 200 μM SAM (*S*-adenosyl-L-methionine) iodide (containing 0.5 μCi of methyl-[H^3]SAM), 1 mM dithiothreitol, and varying concentrations of catechol substrate (e.g., L-dopa, dopamine, or DBA) in a final volume of 1.0 ml. The reaction is initiated by the addition of 250-500 μg of purified DME or crude DME-containing sample and performed at 37 °C for 30 min. The reaction is arrested by rapidly cooling on ice and immediately extracting with 7 ml of ice-cold n-heptane. Following centrifugation at 1000 x g for 10 min, 3-ml aliquots of the organic extracts are analyzed for radioactivity content by liquid scintillation counting. The level of catechol-associated radioactivity in the organic phase is proportional to the catechol-*O*-methyltransferase activity of DME (Zhu, B.T. Liehr, J.G. (1996) 271:1357-1363).

DHFR activity of DME is determined spectrophotometrically at 15 °C by following the disappearance of NADPH at 340 nm (ε_{340} = 11,800 $M^{-1}\cdot cm^{-1}$). The standard assay mixture contains 100 μM NADPH, 14 mM 2-mercaptoethanol, MTEN buffer (50 mM 2-morpholinoethanesulfonic acid, 25 mM tris(hydroxymethyl)aminomethane, 25 mM ethanolamine, and 100 mM NaCl, pH 7.0), and DME in a final volume of 2.0 ml. The reaction is started by the addition of 50 μM dihydrofolate (as substrate). The oxidation of NADPH to NADP$^+$ corresponds to the reduction of dihydrofolate in the reaction and is proportional to the amount of DHFR activity in the sample (Nakamura, T. and Iwakura, M. (1999) J. Biol. Chem. 274:19041-19047).

Aldo/keto reductase activity of DME is measured using the decrease in absorbance at 340 nm as NADPH is consumed. A standard reaction mixture is 135 mM sodium phosphate buffer (pH 6.2-7.2 depending on enzyme), 0.2 mM NADPH, 0.3 M lithium sulfate, 0.5-2.5 μg enzyme and an appropriate level of substrate. The reaction is incubated at 30 °C and the reaction is monitored

continuously with a spectrophotometer. Enzyme activity is calculated as mol NADPH consumed / μg of enzyme.

Alcohol dehydrogenase activity of DME is measured using the increase in absorbance at 340 nm as NAD^+ is reduced to NADH. A standard reaction mixture is 50 mM sodium phosphate, pH 7.5, and 0.25 mM EDTA. The reaction is incubated at 25 °C and monitored using a spectrophotometer. Enzyme activity is calculated as mol NADH produced / μg of enzyme.

Carboxylesterase activity of DME activity is determined using 4-methylumbelliferyl acetate as a substrate. The enzymatic reaction is initiated by adding approximately 10 μl of DME-containing sample to 1 ml of reaction buffer (90 mM KH_2PO_4, 40 mM KCl, pH 7.3) with 0.5 mM 4-methylumbelliferyl acetate at 37 °C. The production of 4-methylumbelliferone is monitored with a spectrophotometer ($\varepsilon_{350} = 12.2\ mM^{-1}\ cm^{-1}$) for 1.5 min. Specific activity is expressed as micromoles of product formed per minute per milligram of protein and corresponds to the activity of DME in the sample (Evgenia, V. et al. (1997) J. Biol. Chem. 272:14769-14775).

In the alternative, the cocaine benzoyl ester hydrolase activity of DME is measured by incubating approximately 0.1 ml of enzyme and 3.3 mM cocaine in reaction buffer (50 mM NaH_2PO_4, pH 7.4) with 1 mM benzamidine, 1 mM EDTA, and 1 mM dithiothreitol at 37 °C. The reaction is incubated for 1 h in a total volume of 0.4 ml then terminated with an equal volume of 5% trichloroacetic acid. 0.1 ml of the internal standard 3,4-dimethylbenzoic acid (10 μg/ml) is added. Precipitated protein is separated by centrifugation at 12,000 × g for 10 min. The supernatant is transferred to a clean tube and extracted twice with 0.4 ml of methylene chloride. The two extracts are combined and dried under a stream of nitrogen. The residue is resuspended in 14% acetonitrile, 250 mM KH_2PO_4, pH 4.0, with 8 μl of diethylamine per 100 ml and injected onto a C18 reverse-phase HPLC column for separation. The column eluate is monitored at 235 nm. DME activity is quantified by comparing peak area ratios of the analyte to the internal standard. A standard curve is generated with benzoic acid standards prepared in a trichloroacetic acid-treated protein matrix (Evgenia, V. et al. (1997) J. Biol. Chem. 272:14769-14775).

In another alternative, DME carboxyl esterase activity against the water-soluble substrate para-nitrophenyl butyric acid is determined by spectrophotometric methods well known to those skilled in the art. In this procedure, the DME-containing samples are diluted with 0.5 M Tris-HCl (pH 7.4 or 8.0) or sodium acetate (pH 5.0) in the presence of 6 mM taurocholate. The assay is initiated by adding a freshly prepared para-nitrophenyl butyric acid solution (100 μg/ml in sodium acetate, pH 5.0). Carboxyl esterase activity is then monitored and compared with control autohydrolysis of the substrate using a spectrophotometer set at 405 nm (Wan, L. et al. (2000) J. Biol. Chem. 275:10041-10046).

Sulfotransferase activity of DME is measured using the incorporation of ^{35}S from [^{35}S]PAPS

into a model substrate such as phenol (Folds, A. and Meek, J. L. (1973) Biochim. Biophys. Acta 327:365-374). An aliquot of enzyme is incubated at 37 °C with 1 mL of 10 mM phosphate buffer, pH 6.4, 50 μM phenol, and 0.4-4.0 μM [^{35}S]PAPS. After sufficient time for 5-20% of the radiolabel to be transferred to the substrate, 0.2 mL of 0.1 M barium acetate is added to precipitate protein and phosphate buffer. Then 0.2 mL of 0.1 M $Ba(OH)_2$ is added, followed by 0.2 mL $ZnSO_4$. The supernatant is cleared by centrifugation, which removes proteins as well as unreacted [^{35}S]PAPS. Radioactivity in the supernatant is measured by scintillation. The enzyme activity is determined from the number of moles of radioactivity in the reaction product.

Heparan sulfate 6-sulfotransferase activity of DME is measured in vitro by incubating a sample containing DME along with 2.5 μmol imidazole HCl (pH 6.8), 3.75 μg of protamine chloride, 25 nmol (as hexosamine) of completely desulfated and N-resulfated heparin, and 50 pmol (about 5 x 10^5 cpm) of [^{35}S] adenosine 3′-phosphate 5′-phosphosulfate (PAPS) in a final reaction volume of 50 μl at 37 °C for 20 min. The reaction is stopped by immersing the reaction tubes in a boiling water bath for 1 min. 0.1 μmol (as glucuronic acid) of chondroitin sulfate A is added to the reaction mixture as a carrier. ^{35}S-Labeled polysaccharides are precipitated with 3 volumes of cold ethanol containing 1.3% potassium acetate and separated completely from unincorporated [^{35}S]PAPS and its degradation products by gel chromatography using desalting columns. One unit of enzyme activity is defined as the amount required to transfer 1 pmol of sulfate/min., determined by the amount of [^{35}S]PAPS incorporated into the precipitated polysaccharides (Habuchi, H.et al. (1995) J. Biol. Chem. 270:4172-4179).

In the alternative, heparan sulfate 6-sulfotransferase activity of DME is measured by extraction and renaturation of enzyme from gels following separation by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Following separation, the gel is washed with buffer (0.05 M Tris-HCl, pH 8.0), cut into 3-5 mm segments and subjected to agitation at 4 °C with 100 μl of the same buffer containing 0.15 M NaCl for 48 h. The eluted enzyme is collected by centrifugation and assayed for the sulfotransferase activity as described above (Habuchi, H.et al. (1995) J. Biol. Chem. 270:4172-4179).

In another alternative, DME sulfotransferase activity is determined by measuring the transfer of [^{35}S]sulfate from [^{35}S]PAPS to an immobilized peptide that represents the N-terminal 15 residues of the mature P-selectin glycoprotein ligand-1 polypeptide to which a C-terminal cysteine residue is added. The peptide spans three potential tyrosine sulfation sites. The peptide is linked via the cysteine residue to iodoacetamide-activated resin at a density of 1.5-3.0 μmol peptide/ml of resin. The enzyme assay is performed by combining 10 μl of peptide-derivitized beads with 2-20 μl of DME-containing sample in 40 mM Pipes (pH 6.8), 0.3 M NaCl, 20 mM $MnCl_2$, 50 mM NaF, 1% Triton X-100, and 1 mM 5'-AMP in a final volume of 130 μl. The assay is initiated by addition of

0.5 μCi of [^{35}S]PAPS (1.7 μM; 1 Ci = 37 GBq). After 30 min at 37°C, the reaction beads are washed with 6 M guanidine at 65°C and the radioactivity incorporated into the beads is determined by liquid scintillation counting. Transfer of [^{35}S]sulfate to the bead-associated peptide is measured to determine the DME activity in the sample. One unit of activity is defined as 1 pmol of product formed per min (Ouyang, Y-B. et al. (1998) Biochemistry 95:2896-2901).

In another alternative, DME sulfotransferase assays are performed using [^{35}S]PAPS as the sulfate donor in a final volume of 30 μl, containing 50 mM Hepes-NaOH (pH 7.0), 250 mM sucrose, 1 mM dithiothreitol, 14 μM[^{35}S]PAPS (15 Ci/mmol), and dopamine (25 μM), *p*-nitrophenol (5 μM), or other candidate substrates. Assay reactions are started by the addition of a purified DME enzyme preparation or a sample containing DME activity, allowed to proceed for 15 min at 37 °C, and terminated by heating at 100 °C for 3 min. The precipitates formed are cleared by centrifugation. The supernatants are then subjected to the analysis of ^{35}S-sulfated product by either thin-layer chromatography or a two-dimensional thin layer separation procedure. Appropriate standards are run in parallel with the supernatants to allow the identification of the ^{35}S-sulfated products and determine the enzyme specificity of the DME-containing samples based on relative rates of migration of reaction products (Sakakibara, Y. et al. (1998) J. Biol. Chem. 273:6242-6247).

Squalene epoxidase activity of DME is assayed in a mixture comprising purified DME (or a crude mixture comprising DME), 20 mM Tris-HCl (pH 7.5), 0.01 mM FAD, 0.2 unit of NADPH-cytochrome C (P-450) reductase, 0.01 mM [^{14}C]squalene (dispersed with the aid of 20 μl of Tween 80), and 0.2% Triton X-100. 1 mM NADPH is added to initiate the reaction followed by incubation at 37 °C for 30 min. The nonsaponifiable lipids are analyzed by silica gel TLC developed with ethyl acetate/benzene (0.5:99.5, v/v). The reaction products are compared to those from a reaction mixture without DME. The presence of 2,3(*S*)-oxidosqualene is confirmed using appropriate lipid standards (Sakakibara, J. et al. (1995) 270:17-20).

Epoxide hydrolase activity of DME is determined by following substrate depletion using gas chromatographic (GC) analysis of ethereal extracts or by following substrate depletion and diol production by GC analysis of reaction mixtures quenched in acetone. A sample containing DME or an epoxide hydrolase control sample is incubated in 10 mM Tris-HCl (pH 8.0), 1 mM ethylenediaminetetraacetate (EDTA), and 5 mM epoxide substrate (e.g., ethylene oxide, styrene oxide, propylene oxide, isoprene monoxide, epichlorohydrin, epibromohydrin, epifluorohydrin, glycidol, 1,2-epoxybutane, 1,2-epoxyhexane, or 1,2-epoxyoctane). A portion of the sample is withdrawn from the reaction mixture at various time points, and added to 1 ml of ice-cold acetone containing an internal standard for GC analysis (e.g., 1-nonanol). Protein and salts are removed by centrifugation (15 min, 4000 × g) and the extract is analyzed by GC using a 0.2 mm × 25-m CP-Wax57-CB column (CHROMPACK, Middelburg, The Netherlands) and a flame-ionization

detector. The identification of GC products is performed using appropriate standards and controls well known to those skilled in the art. 1 Unit of DME activity is defined as the amount of enzyme that catalyzes the production of 1 μmol of diol/min (Rink, R. et al. (1997) J. Biol. Chem. 272:14650-14657).

Aminotransferase activity of DME is assayed by incubating samples containing DME for 1 hour at 37 °C in the presence of 1 mM L-kynurenine and 1 mM 2-oxoglutarate in a final volume of 200 μl of 150 mM Tris acetate buffer (pH 8.0) containing 70 μM PLP. The formation of kynurenic acid is quantified by HPLC with spectrophotometric detection at 330 nm using the appropriate standards and controls well known to those skilled in the art. In the alternative, L-3-hydroxykynurenine is used as substrate and the production of xanthurenic acid is determined by HPLC analysis of the products with UV detection at 340 nm. The production of kynurenic acid and xanthurenic acid, respectively, is indicative of aminotransferase activity (Buchli, R. et al. (1995) J. Biol. Chem. 270:29330-29335).

In another alternative, aminotransferase activity of DME is measured by determining the activity of purified DME or crude samples containing DME toward various amino and oxo acid substrates under single turnover conditions by monitoring the changes in the UV/VIS absorption spectrum of the enzyme-bound cofactor, pyridoxal 5'-phosphate (PLP). The reactions are performed at 25 °C in 50 mM 4-methylmorpholine (pH 7.5) containing 9 μM purified DME or DME containing samples and substrate to be tested (amino and oxo acid substrates). The half-reaction from amino acid to oxo acid is followed by measuring the decrease in absorbance at 360 nm and the increase in absorbance at 330 nm due to the conversion of enzyme-bound PLP to pyridoxamine 5' phosphate (PMP). The specificity and relative activity of DME is determined by the activity of the enzyme preparation against specific substrates (Vacca, R.A. et al. (1997) J. Biol. Chem. 272:21932-21937).

Superoxide dismutase activity of DME is assayed from cell pellets, culture supernatants, or purified protein preparations. Samples or lysates are resolved by electrophoresis on 15% non-denaturing polyacrylamide gels. The gels are incubated for 30 min in 2.5 mM nitro blue tetrazolium, followed by incubation for 20 min in 30 mM potassium phosphate, 30 mM TEMED, and 30 μM riboflavin (pH 7.8). Superoxide dismutase activity is visualized as white bands against a blue background, following illumination of the gels on a lightbox. Quantitation of superoxide dismutase activity is performed by densitometric scanning of the activity gels using the appropriate superoxide dismutase positive and negative controls (e.g., various amounts of commercially available E. coli superoxide dismutase (Harth, G. and Horwitz, M.A. (1999) J. Biol. Chem. 274:4281-4292).

XVIII. Identification of DME Inhibitors

Compounds to be tested are arrayed in the wells of a multi-well plate in varying

concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVII. DME activity is measured for each well and the ability of each compound to inhibit DME activity can be determined, as well as the dose-response profiles. This assay could also be used to identify molecules which enhance DME activity.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Nucleotide SEQ ID NO:	Incyte Nucleotide ID
1799250	1	1799250CD1	25	1799250CB1
2242475	2	2242475CD1	26	2242475CB1
2706492	3	2706492CD1	27	2706492CB1
2766688	4	2766688CD1	28	2766688CB1
2788823	5	2788823CD1	29	2788823CB1
3348822	6	3348822CD1	30	3348822CB1
4290251	7	4290251CD1	31	4290251CB1
4904188	8	4904188CD1	32	4904188CB1
638419	9	638419CD1	33	638419CB1
1844394	10	1844394CD1	34	1844394CB1
2613056	11	2613056CD1	35	2613056CB1
5053617	12	5053617CD1	36	5053617CB1
5483256	13	5483256CD1	37	5483256CB1
5741354	14	5741354CD1	38	5741354CB1
5872615	15	5872615CD1	39	5872615CB1
2657543	16	2657543CD1	40	2657543CB1
3041639	17	3041639CD1	41	3041639CB1
3595451	18	3595451CD1	42	3595451CB1
4169101	19	4169101CD1	43	4169101CB1
2925182	20	2925182CD1	44	2925182CB1
3271838	21	3271838CD1	45	3271838CB1
3292871	22	3292871CD1	46	3292871CB1
4109179	23	4109179CD1	47	4109179CB1
4780365	24	4780365CD1	48	4780365CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
1	1799250CD1	g9622124	7.00e-39	androgen-regulated short-chain dehydrogenase/reductase 1 [Homo sapiens]
2	2242475CD1	g181350	2.3e-250	debrisoquine 4-hydroxylase [Homo sapiens]
3	2706492CD1	g9622124	2.00e-29	androgen-regulated short-chain dehydrogenase/reductase 1 [Homo sapiens]
4	2766688CD1	g7533022	0	oxysterol 7alpha-hydroxylase [Mus musculus]
5	2788823CD1	g9622124	4.00e-63	androgen-regulated short-chain dehydrogenase/reductase 1 [Homo sapiens]
6	3348822CD1	g164981	7.9e-138	cytochrome P-450p-2 [Oryctolagus cuniculus]
7	4290251CD1	g3135970	2.7e-150	dJ352A20.2 (aldehyde dehydrogenase family)
8	4904188CD1	g10039619	2.00e-92	PAN2 [Homo sapiens]
9	638419CD1	g3004922	3.8e-99	phenol sulfotransferase [Mus musculus]
10	1844394CD1	g2621120	7.6e-14	O-linked GlcNAc transferase [Methanobacterium themoautotrophicum]
11	2613056CD1	g6063487	7.1e-100	cytochrome P450 XL-301 [Xenopus laevis]
12	5053617CD1	g2662573	0.00075	similar to UDP-glucuronosyltransferase [Caenorhabditis elegans]
13	5483256CD1	g3879119	5.5e-61	similar to Glutathione S-transferases.; cDNA EST...
14	5741354CD1	g1185452	1.5e-104	cytochrome P450 monooxygenase CYP2J2 [Homo sapiens]
15	5872615CD1	g5410280	7.0e-38	HSPC034 protein [Homo sapiens]
		g2921821	0.81	Cytochrome P450 IIE1 [Rattus norvegicus]

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
16	2657543CD1	g510905	4.5e-27	glutathione transferase T1 [Homo sapiens]
17	3041639CD1	g1280387	5.3e-95	alpha 2,6-sialyltransferase [Rattus norvegicus]
18	3595451CD1	g3355516	1.6e-118	dJ248E1.1 (DOPAMINE-BETA-MONOOXYGENASE PRECURSOR (DOPAMINE BETA-HYDROXYLASE) (DBH)) [Homo sapiens]
19	4169101CD1	g1055177	6.0e-97	weakly similar to E. nidulans bimA gene product
		g2621120	1.1e-28	O-linked GlcNAc transferase [Methanobacterium thermoautotrophicum]
20	2925182CD1	g6329074	5.0e-289	UDP-GlcNAc:a-1,3-D-mannoside b-1,4-N-acetylglucosaminyltransferase IV [Homo sapiens]
21	3271838CD1	g4827177	2.6e-285	thioredoxin reductase II alpha [Homo sapiens]
22	3292871CD1	g8515441	0	cytochrome P450 retinoid metabolizing protein P450RAI-2 [Homo sapiens]
23	4109179CD1	g155947	2.4e-06	cytochrome P450 [Blaberus discoidalis]
24	4780365CD1	g4590450	2.9e-183	A1-specific alpha 1->3 N-acetylgalactosaminyltransferase [Homo sapiens]

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino acid residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences and Motifs	Analytical Methods and Databases
1	1799250CD1	330	T59 T91 S108 T165 S175 S269	N76	Signal Peptide:M1-G31 Short-chain alcohol dehydrogenase:R44-D246, D195-L222, I121-G131, V188-L225, H232-G241 Short-chain alcohol dehydrogenase family: DM00034|S42651|28-318:A46-M322	BLAST-DOMO BLIMPS-BLOCKS BLIMPS-PRINTS HMMER Motifs SPScan
2	2242475CD1	497	T93 T138 S168 T249 S289 T378 S379 T407	N166 N398	Signal peptide: M1-R25 Transmembrane domain:M1-L18 p450:P34-A494 Cytochrome P450 cysteine BL00086:F422-F453 E-class P450 group I sig: PR00463A-I:R62-L81, A86-T107, A181-D199, N294-S311, L314-G340, D346-Q364, F387-K411, D422-A432, A432-L455 Cytochrome_P450: P34-A494, F404-S454, F436-L444 Cytochrome P450 heme-iron ligand sig: H416-S465 Cytochrome P450 family: DM00022|I49427|46-483:L43-F481 PD000021:L251-G386	BLAST-DOMO BLAST-PRODOM BLIMPS HMMER Motifs ProfileScan
3	2706492CD1	286	S10 S187 S216 S264 S265 T272		Signal cleavage:M1-A68 Short-chain alcohol dehydrogenase family: M1-E142, BL00061A:P23-G33, BL00061C:G128-G137, PR00080A:P23-I34, PR00080C:Y106-E125 Short-chain alcohol dehydrogenase family: DM00034|S42651|28-318:M1-L219	BLAST-DOMO BLIMPS-BLOCKS BLIMPS-PRINTS Motifs SPScan

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino acid residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences and Motifs	Analytical Methods and Databases
4	2766688CD1	469	T77 S88 T182 S218 S238 T258 S259 S291 S318 T335 T345 S409 S438	N176 N214	Signal peptide:M1-Q22 Transmembrane domain:S5-Q22 Cytochrome P450:P29-R396 E-class cytochrome P450 sig: PR00465A-F,H:P29-G46, E51-T74, P264-L290, L325-P341, Y357-W371, H373-E391, C414-L432 Cytochrome P450: DM02967|Q09736|16-486:L21-W390 DM00022|S50211|59-488:W272-E458	BLAST-DOMO BLIMPS-PRINTS HMMER Motifs
5	2788823CD1	331	T54 S100 S103 T134 T135 T191 S215 S284 S313 S323	N171	Signal cleavage:M1-A17 Short-chain dehydrogenase:K39-E236 Short-chain dehydrogenase: BL00051A-C:E116-G126, G180-R217, G222-G231 Alcohol dehydrogenase: PR00080A-C:E116-V127, S167-L175, Y200-Q219 Short-chain alcohol dehydrogenase family: DM00034|S42651|28-318:T40-V317	BLAST-DOMO BLIMPS-BLOCKS BLIMPS-PRINTS Motifs SPScan
6	3348822CD1	509	S4 S104 T106 S159 S172 T173 T176 S177 S207 T208 T278 S292 S300 S302 T374 T393	N206	Signal peptide:M1-L32 Signal cleavage:M1-A29 Transmembrane domain:L16-L34 Cytochrome P450:F46-L503 Cytochrome P450 cys heme-iron ligand: F426-H474 Cytochrome P450 cysteine: BL00086: Y444-F475 Cytochrome P450: DM00022|P10611|120-497:P119-M499	BLAST-DOMO BLIMPS-BLOCKS HMMER Motifs SPScan Profilescan

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino acid residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences and Motifs	Analytical Methods and Databases
7	4290251CD1	433	S23 Y24 S31 T32 S42 S62 S63 S65 S83 T129 T140 T162 S220 T275 S350 T430		Aldehyde dehydrogenase family: aldedh:K17-H433 aldehyde_dehydr_glu.prf:G211-S279 Aldehyde dehydrogenase: L242-P259 Aldoketoreductase 3: L101-F116 Aldehyde dehydrogenase: BL00687A-F: W61-D78, A143-S184, P200-A236, G255-G301, G316-L365, P402-G412 Aldehyde dehydrogenase family: PD000218:D29-L307 DM00100|P19059|1-462:L10-Y317	BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS Motifs HMMER Motifs ProfileScan
8	4904188CD1	186	S41 S42 T128 S141 S142 S149	N62 N89	Signal cleavage:M1-S31 Glucose rubitol dehydrogenase: PR00081C-F: L29-Y45, Y67-E86, T88-G105, W122-S142 Short-chain alcohol dehydrogenase family: DM00034|Q03326|1-259:D5-E158 DM00034|S39394|69-356:T3-S42	BLAST-DOMO BLIMPS-PRINTS Motifs SPScan
9	638419CD1	304	S173 S180 T194 T236 S237	N248 N258	Sulfotransferase: PD001218:F18-K292 PAPS Cofactor Binding Site: DM00981|P52840|1-291:I20-I304	BLAST-DOMO BLAST-PRODOM HMMER Motifs
10	1844394CD1	629	S11 S83 T92 T148 T150 T179 S185 T186 S209 S299 Y310 S312 S339 S366 T403 T424 T466 T488 S510 S542 S566 S576 T609 T284	N386 N563	TPR Domain TPR:Q316-P344, W350-P378	HMMER-PFAM
					HYPOTHETICAL 90.0 KD PROTEIN T20B12.1 IN CHROMOSOME III PD141851:K2-L246	BLAST-PRODOM
					Intermediate filaments signature I435-D443	MOTIFS
					Cytochrome c and c1 heme BL00821E:L71-N84	BLIMPS-BLOCKS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino acid residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences and Motifs	Analytical Methods and Databases
11	2613056CD1	320	S21 T72 S82 T92 S105 S115 T206 S249	N32 N196	Cytochrome P450 p450:M1-A316	HMMER-PFAM
					Cytochrome_P450 F260-G269	MOTIFS
					Cytochrome P450 cysteine heme-iron ligand signature cytochrome_p450.prf:F239-R287	ProfileScan
					CYTOCHROME P450 MONOOXYGENASE OXIDOREDUCTASE HEME ELECTRON TRANSPORT MEMBRANE MICROSOME ENDOPLASMIC PD000021:L50-S198	BLAST-PRODOM
					CYTOCHROME P450 DM00022|P10611|120-497:M1-I312	BLAST-DOMO
					Cytochrome P450 cysteine heme-iron ligand signature BL00086:F257-F288	BLIMPS-BLOCKS
					E-class P450 group II signature PR00464B:M1-Q18 PR00464C:D118-A146, PR00464D:K147-G164, PR00464E:Q176-N196, PR00464F:G216-Y231, PR00464G:F232-E247, PR00464H:P254-C267, PR00464I:C267-L290	BLIMPS-PRINTS
12	5053617CD1	56				
13	5483256CD1	377	S55 T75 S97 S144 T253 S290	N163	signal_cleavage:M1-A50	SPSCAN
					Glutathione S-transferases GST:L102-V152, A278-I370	HMMER-PFAM
					Glutaredoxin proteins: BL00195A:L104-V116	BLIMPS-BLOCKS
					Glutaredoxin signature: PR00160A:L102-L120	BLIMPS-PRINTS
					PROTEIN SUP R11A8.5: PD134628:S98-E372	BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino acid residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences and Motifs	Analytical Methods and Databases
14	5741354CD1	501	S70 S143 T171 T201 S303 S351 T383 Y431	N189	CYTOCHROME P450 DM00022|P52786|83-492:G83-P491	BLAST-DOMO
					Cytochrome_P450: F441-L449	MOTIFS
					signal_peptide:M1-L31	HMMER
					signal_cleavage:M1-Q29	SPSCAN
					transmem_domain:A12-L30	HMMER
					Cytochrome P450: p450:P40-A498	HMMER-PFAM
					Cytochrome P450 cysteine heme-iron ligand signature: BL00086:L438-F469	BLIMPS-BLOCKS
					E-class P450 group I signature PR00463A:S70-L89, PR00463B:V94-F115, PR00463C:A186-D204, PR00463D:N299-T316, PR00463E:L319-G345, PR00463F:E362-F380, PR00463G:N403-D427, PR00463H:L438-C448, PR00463I:C448-L471	BLIMPS-PRINTS
					Cytochrome P450 cysteine heme-iron ligand signature: cytochrome_p450.prf:F420-H470	ProfileScan
					CYTOCHROME P450 MONOOXYGENASE OXIDOREDUCTASE HEME ELECTRON TRANSPORT MEMBRANE MICROSOME ENDOPLASMIC PD000021:Q244-G391	BLAST-PRODOM
15	5872615CD1	144	T19 S20 T71 T78 S121	N141		
16	2657543CD1	218	T90 S72 S152		signal peptide:M1-G23	SPScan
					GST (glutathione S-transferases):N9-V161	HMMER-PFAM
					Dichloromethane dehalogenase: DM02033|Q01579|70-200:V68-M162	BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation sites	Signature Sequences and Motifs	Analytical Methods and Databases
17	3041639CD1	210	S9 T53 T189 T204 S194	N148	signal_cleavage:M1-E30	SPScan
					ALPHANACETYLGALACTOSAMINIDE ALPHA2,6SIALYLTRANSFERASE: PD129519:M1-Q74	BLAST-PRODOM
					GLYCOSYLTRANSFERASE: PD129520:W155-D207	BLAST-PRODOM
					LUMENAL DOMAIN, SIALYLTRANSFERASE: DM01020|S41114|71-400:C80-P157	BLAST-DOMO
18	3595451CD1	613	S21 T29 S42 T118 S135 T140 S162 S168 S249 S253 T295 T312 S459 T460 S482 S511 S519 T530 S551 S609	N114 N247 N476 N517	COPPER TYPE II, ASCORBATE-DEPENDENT MONOOXYGENASES: DM04634|P08478|1-399:T268-Y472	BLAST-DOMO
					PEPTIDYLGLYCINE MONOOXYGENASE I: DM07918|P15101|1-609:L6-P573	BLAST-DOMO
					signal_peptide:M1-A18	HMMER
					signal_cleavage:M1-A18	SPScan
					Copper type II, ascorbate-dependent monooxygenase: Cu2_monooxygen:M1-G337	HMMER-PFAM
					Copper type II, ascorbate-dependent monooxygenases signature: cu2_monooxygenase_1.prf:V212-L317	ProfileScan
					DOPAMINE BETAMONOOXYGENASE: PD014255:V338-P573	BLAST-PRODOM
					MONOOXYGENASE: PD004410:D193-W344	BLAST-PRODOM
					Dopamine-beta-monooxygenase: PR00767C:I120-D138, PR00767D:Y203-E223, PR00767E:V225-S243, PR00767G:V272-L291, PR00767H:A336-P353	BLIMPS-PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation sites	Signature Sequences and Motifs	Analytical Methods and Databases
19	4169101CD1	741	T6 S80 S83 Y84 T91 T136 S195 T299 S306 S349 T436 T454 S562 S572 Y588 S665	N78 N497 N609	Aldoketo_Reductase_3: L346-L361	Motifs
					transmem_domain:I112-F130	HMMER
					transmem_domain:K440-R460	HMMER
					TPR Domain (tetratricopeptide repeat): TPR (6 domains): H485-P513, M519-P547, Y587-P615, W621-P649, M655-P683, H689-P717	HMMER-PFAM
					F32D1.3 PROTEIN SIMILAR E NIDULANS BIMA GENE PRODUCT: PD041324:Y285-L447	BLAST-PRODOM
20	2925182CD1	535	S56 T90 T98 S99 T143 S158 T180 S209 T224 T272 S281 S369 T376 Y380 S399 T459 Y486 T534	N5 N77 N458	signal_peptide:M1-S22	HMMER
					signal_cleavage:M1-Y24	SPScan
					transmem_domain:S282-M299	HMMER
					UDPGLCNAC:A1,3D MANNOSIDE B1,4NACETYLGLUCOSAMINYLTRANSFERASE IV EC 2.4.1.145: PD185013:M1-N535	BLAST-PRODOM
21	3271838CD1	522	S66 S160 T196 T207 T295 T301 T334 T338 T401 S411 T499 S510 S513		Pyridine_Redox_1:G83-P93	Motifs
					Pyridine nucleotide-disulphide oxidoreductase: pyr_redox:L42-V519	HMMER-PFAM
					Pyridine nucleotide-disulphide oxidoreductases class-I active site: pyridine_redox_1.prf:C54-Y113	ProfileScan
					Pyridine nucleotide-disulphide oxidoreductases class-I: BL00076A:Y40-D69, BL00076B:G83-K95 BL00076C:D310-T349	BLIMPS-BLOCKS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation sites	Signature Sequences and Motifs	Analytical Methods and Databases
21 (cont)					Pyridine nucleotide disulfide oxidoreductase: PR00411A:D41-R63, PR00411B:L82-M97, PR00411C:I184-R193, PR00411D:K220-T245, PR00411E:D310-S324, PR00411F:I354-V361, PR00411G:D391-E412, PR00411H:V460-Q475, PR00411I:K482-E502	BLIMPS-PRINTS
					FAD-dependent pyridine n...: PR00368A:D41-R63, PR00368B:I184-R193, PR00368C:K220-T245, PR00368D:D310-S324, PR00368E:I354-V361	BLIMPS-PRINTS
					FAD REDUCTASE REDOXACTIVE CENTER: PD000139:V294-K482	BLAST-PRODOM
					PYRIDINE NUCLEOTIDE-DISULPHIDE OXIDOREDUCTASES CLASS-I: DM00071|S57658|39-320:D39-D321	BLAST-DOMO
22	3292871CD1	495	S32 T41 S57 S58 T80 S103 S116 T151 S157 T172 S184 S256 T265 T300 T325 S378 T444 S445 S470	N3	Cytochrome_P450: F417-G426	Motifs
					signal_cleavage:M1-G19	SPScan
					Cytochrome P450: W48-L89, E160-L432	HMMER-PFAM
					Cytochrome P450 cysteine heme-iron ligand signature: D396-L441	ProfileScan
					Cytochrome P450 cysteine heme-iron ligand proteins: BL00086:Y414-S445	BLIMPS-BLOCKS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation sites	Signature Sequences and Motifs	Analytical Methods and Databases
22 (cont)					Mitochondrial P450 signature: PR00408E:K297-R310, PR00408F:S332-P350, PR00408H:L415-C424, PR00408I:C424-K435	BLIMPS-PRINTS
					P450 superfamily signature: PR00385A:A279-L296, PR00385B:K297-R310, PR00385C:C339-P350, PR00385D:L415-C424, PR00385E:C424-K435	BLIMPS-PRINTS
					CYTOCHROME P450: DM00022|P08684|58-487:Q221-P465	BLAST-DOMO
23	4109179CD1	51			Cytochrome P450: M1-R45 (Score=9.6, E-value=0.027)	HMMER-PFAM
					Cytochrome P450: DM00022|P29981|79-497:M1-I39 (p=1.2e-6)	BLAST-DOMO
24	4780365CD1	335	S19 T147 S152 S166 S183 Y187 S220 Y246 S254	N94	signal_peptide:M1-S19	HMMER
					signal_cleavage:M1-G14	SPScan
					transmem_domain:M1-P20	HMMER
					GALACTOSYLTRANSFERASE: PD010022:L47-N334	BLAST-PRODOM
					TRANSFERASE HISTOBLOOD GROUP ABO SYSTEM: PD041469:M1-L47	BLAST-PRODOM
					SIGNAL-ANCHOR TRANSMEM (includes galactosyltransferases): DM07533|P16442|16-353:M1-P335	BLAST-DOMO
					GALACTOSYLTRANSFERASE: DM08008|I49698|1-371:R62-N334	BLAST-DOMO

Table 4

Nucleotide SEQ ID NO:	Incyte Nucleotide ID	Sequence length	Selected fragment(s)	Sequence fragments	5' position	3' position
25	1799250CB1	1269	1-630	6537030H1 (OVARDIN02)	580	1169
				1620357T6 (BRAITUT13)	706	1269
				1607327H1 (LUNGNOT15)	1	205
				1620357F6 (BRAITUT13)	16	502
				1799250F6 (COLNNOT27)	218	663
				6846328H1 (KIDNTMN03)	689	1223
26	2242475CB1	1593	323-848, 1-116	3699419H1 (SININOT05)	1	293
				70614021V1	446	1165
				70614588V1	1022	1622
				70611772V1	5	645
27	2706492CB1	1779	1-1058	1921856R6 (BRSTTUT01)	1291	1779
				6969174U1	465	1204
				2706492F6 (PONSAZT01)	1	554
				6568102H1 (MCLDTXN05)	914	1433
28	2766688CB1	1931	1-1041	5893501H1 (BRAYDIN03)	1112	1340
				4273116H1 (PROSTMT01)	899	1183
				g758933	1221	1831
				6702946H1 (DRGCNOT02)	115	751
				2950040H1 (KIDNFET01)	1	294
				2514733F7 (LIVRTUT04)	367	902
				6336624H1 (BRANDIN01)	1382	1931
				2766688F6 (BRSTNOT12)	498	1010
29	2788823CB1	1282	1-399, 1252-1282, 427-529	6875661H1 (EPIMUNN04)	85	643
				6333929H1 (BRANDIN01)	665	1141
				1944989H1 (PITUNOT01)	1045	1282
				6916176H1 (PLACFER06)	336	919
				2788823H1 (HEAONOT02)	32	315
				1573860H1 (LNODNOT03)	1	26

Table 4 (cont.)

Nucleotide SEQ ID NO:	Incyte Nucleotide ID	Sequence length	Selected fragment(s)	Sequence fragments	5' position	3' position
30	3348822CB1	2416	2339-2416, 1-1024, 1515-1636	70882367V1	945	1462
				6620514H1 (MUSLTDR02)	851	1390
				6574030H1 (COLHTUS02)	1328	1918
				2525166H1 (BRAITUT21)	1	243
				60138453V1	2052	2338
				70879864V1	1477	2103
				7176308H1 (BRSTTMC01)	357	953
				1985721H1 (LUNGAST01)	2162	2416
				6420342H1 (BRSTUNT01)	126	679
31	4290251CB1	1574	1-29, 1359-1574, 859-997, 402-454	2515438F6 (LIVRTUT04)	38	627
				4290037R6 (BRABDIR01)	1081	1574
				5521434H1 (LIVRDIR01)	1	266
				4701356H1 (SMCRTXT01)	956	1223
				4290251F6 (BRABDIR01)	554	1080
32	4904188CB1	1227	1-403	5543059T8 (TESTNOC01)	283	1212
				70327902D1	1	551
				70326347D1	785	1227
				4029455T6 (BRAINOT23)	603	1224
33	638419CB1	1240	1-496, 1028-1240	638419H1 (BRSTNOT03)	1011	1240
				70568268V1	1	600
				2817590T6 (BRSTNOT14)	726	1221
				70568923V1	439	1034

Table 4 (cont.)

Nucleotide SEQ ID NO:	Incyte Nucleotide ID	Sequence length	Selected fragment(s)	Sequence fragments	5' position	3' position
34	1844394	2275	1-431, 1030-1366	2501918H1 (ADRETUT05)	1	240
				2501918F6 (ADRETUT05)	1	543
				2669184F6 (ESOGTUT02)	802	1464
				3114958H1 (BRSTNOT17)	283	610
				3135773F6 (SMCCNOT01)	1898	2275
				2343269F6 (TESTTUT02)	617	1235
				4529167H1 (LYMBTXT01)	1311	1601
				5571305H1 (TLYMNOT08)	367	622
				2669184T6 (ESOGTUT02)	1608	2247
				2057708T6 (BEPINOT01)	1316	2246
35	2613056	1586	400-532	2742715F6 (BRSTTUT14)	490	1025
				2743015X306D1(BRSTTUT14)	1	487
				4187166F6 (BRSTNOT31)	449	936
				2742715T6 (BRSTTUT14)	930	1586
36	5053617	859	1-291	g1330846	275	859
				206675F1 (SPLNNOT02)	422	837
				001071H1 (U937NOT01)	480	853
				206675R1 (SPLNNOT02)	313	831
				3880457F6 (SPLNNOT11)	1	343
37	5483256	2302	1-1336, 2022-2302	6144620H1 (BRANDIT03)	431	1075
				7179844H1 (BRAXDIC01)	1	574
				6773457J1 (OVARDIR01)	1730	2302
				930526R1 (CERVNOT01)	769	1326
				2059641T6 (OVARNOT03)	1350	2009
				6576653H1 (COLHTUS02)	1284	1982
38	5741354	1653	1-985	2229885F6 (PROSNOT16)	800	1280
				2717950F6 (THYRNOT09)	1	618
				g1551782	1077	1653
				494427F1 (HNT2NOT01)	1104	1653
				SBLA03737F1	496	1049

Table 4 (cont.)

Nucleotide SEQ ID NO:	Incyte Nucleotide ID	Sequence length	Selected fragment(s)	Sequence fragments	5' position	3' position
39	5872615	683	1-25	6873589H1 (EPIMUNN04)	47	683
				2938376H1 (THYMFET02)	1	260
40	2657543	657	161-181, 343-657	g5420326.v113.gs_3.nt	1	657
				2657543H1 (LUNGTUT09)	341	568
41	3041639	1122	637-808, 1-138, 911-1122	70827651V1	435	1025
				3041639T6 (BRSTNOT16)	466	1096
				70827016V1	1	470
42	3595451	2982	1-2274	70465056V1	1759	2410
				70465685V1	1136	1753
				3595451F6 (FIBPNOT01)	123	686
				70465630V1	2330	2982
				70466466V1	1024	1692
				7255550H1 (FIBRTXC01)	459	1116
				3596949H1 (FIBPNOT01)	1	300
				70467798V1	1705	2399
43	4169101	3517	3409-3517, 1-1773	70484641V1	1373	2025
				6849059H1 (KIDNTMN03)	2060	2633
				70483292V1	814	1408
				6757006J1 (SINTFER02)	1	711
				1546352H1 (PROSTUT04)	3316	3517
				6494614H1 (BONRNOT01)	2692	3344
				6937377H1 (FTUBTUR01)	1623	2128
				70483348V1	663	1173
				1691757T6 (PROSTUT10)	2737	3379
				2499325F6 (ADRETUT05)	2217	2716

Table 4 (cont.)

Nucleotide SEQ ID NO:	Incyte Nucleotide ID	Sequence length	Selected fragment(s)	Sequence fragments	5' position	3' position
44	2925182	2339	368-920, 1-20	3423537H1 (BRSTNOR01)	510	750
				7094242H1 (SINITME01)	1378	1947
				6595856H1 (COLENOR03)	743	1371
				4447810H1 (SINDNOT01)	407	666
				60141115D1	1	462
				70684441V1	1734	2339
				60141117D1	624	1078
				4190070T6 (BRAPDIT01)	1347	1931
45	3271838	1955	1-943	2519257H1 (BRAITUT21)	1	253
				2101033R6 (BRAITUT02)	499	1008
				60213368U1	1027	1443
				7178847H1 (BRAXDIC01)	102	693
				1998271R6 (BRSTTUT03)	1425	1955
				2417893F6 (HNT3AZT01)	896	1378
46	3292871	2065	791-1046, 1-159, 1481-2065, 258-292	g5091644.v113.gs_1.1.nt.edit	147	410
				3537160H1 (SEMVNOT04)	1	203
				3292871F6 (BONRFET01)	1055	1462
				6754650J1 (SINTFER02)	260	902
				6314941H1 (NERDTDN03)	1435	2065
				6819509H1 (OVARDIR01)	609	1217
47	4109179	866	1-87, 697-866	7267351H1 (NOSEDIC01)	1	523
				5872309H1 (COLTDIT04)	585	866
				7093802H1 (SINITME01)	548	800
				6847686H1 (KIDNTMN03)	370	672
48	4780365	1593	791-1048, 1-717	6801412J1 (COLENOR03)	347	834
				6755595J1 (SINTFER02)	1	424
				6829516H1 (SINTNOR01)	1173	1593
				6801420H1 (COLENOR03)	718	1196

Table 5

Nucleotide SEQ ID NO:	Incyte Project ID	Representative Library
25	1799250	TESTNOT17
26	2242475	BRSTNON02
27	2706492	COLNNOT19
28	2766688	PROSTMT01
29	2788823	OVARDIN02
30	3348822	BRAITUT21
31	4290251	BRABDIR01
32	4904188	SININOT04
33	638419	BRSTNOT14
34	1844394	BEPINOT01
35	2613056	BRSTTUT14
36	5053617	URETTUT01
37	5483256	BRAINOT09
38	5741354	THYRNOT09
39	5872615	CONUTUT01
40	2657543	LUNGTUT09
41	3041639	BONSTUT01
42	3595451	KIDNNOT05
43	4169101	FIBPFEN06
44	2925182	SINITME01
45	3271838	THP1AZT01
46	3292871	BONRFET01
47	4109179	PROSBPT07
48	4780365	SINTNOR01

Table 6

Library	Vector	Library Description
BRABDIR01	pINCY	Library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident. Patient history included Huntington's disease, emphysema, and tobacco abuse.
BRAITUT21	pINCY	Library was constructed using RNA isolated from brain tumor tissue removed from the midline frontal lobe of a 61-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated subfrontal meningothelial meningioma with no atypia. One ethmoid and mucosal tissue sample indicated meningioma. Family history included cerebrovascular disease, senile dementia, hyperlipidemia, benign hypertension, atherosclerotic coronary artery disease, congestive heart failure, and breast cancer.
BRSTNON02	pINCY	This normalized breast tissue library was constructed from 6.2 million independent clones from a pool of two libraries from two different donors. Starting RNA was made from breast tissue removed from a 46-year-old Caucasian female during a bilateral reduction mammoplasty (donor A), and from breast tissue removed from a 60-year-old Caucasian female during a bilateral reduction mammoplasty (donor B). Pathology indicated normal breast parenchyma, bilaterally (A) and bilateral mammary hypertrophy (B). Patient history included hypertrophy of breast, obesity, lumbago, and glaucoma (A) and joint pain in the shoulder, thyroid cyst, colon cancer, normal delivery and cervical cancer (B). Family history included cataract, osteoarthritis, uterine cancer, benign hypertension, hyperlipidemia, and alcoholic cirrhosis of the liver, cerebrovascular disease, and type II diabetes (A) and cerebrovascular accident, atherosclerotic coronary artery disease, colon cancer, type II diabetes, hyperlipidemia, depressive disorder, and Alzheimer's Disease. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.

Table 6 (cont.)

Library	Vector	Library Description
BRSTNOT14	pINCY	Library was constructed using RNA isolated from breast tissue removed from a 62-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated an invasive grade 3 (of 4), nuclear grade 3 (of 3) adenocarcinoma, ductal type. Ductal carcinoma in situ, comedo type, comprised 60% of the tumor mass. Metastatic adenocarcinoma was identified in one (of 14) axillary lymph nodes with no perinodal extension. The tumor cells were strongly positive for estrogen receptors and weakly positive for progesterone receptors. Patient history included a benign colon neoplasm, hyperlipidemia, cardiac dysrhythmia, and obesity. Family history included atherosclerotic coronary artery disease, myocardial infarction, colon cancer, ovarian cancer, lung cancer, and cerebrovascular disease.
COLNNOT19	pINCY	Library was constructed using RNA isolated from the cecal tissue of an 18-year-old Caucasian female. The cecal tissue, along with the appendix and ileum tissue, were removed during bowel anastomosis. Pathology indicated Crohn's disease of the ileum, involving 15 cm of the small bowel.
OVARDIN02	pINCY	This normalized ovarian tissue library was constructed from 5.76 million independent clones from an ovary library. Starting RNA was made from diseased ovarian tissue removed from a 39-year-old Caucasian female during total abdominal hysterectomy, bilateral salpingo-oophorectomy, dilation andcurettage, partial colectomy, incidental appendectomy, and temporary colostomy. Pathology indicated the right and left adnexa, mesentery and muscularis propria of the sigmoid colon were extensively involved by endometriosis. Endometriosis also involved the anterior and posterior serosal surfaces of the uterus and the cul-de-sac. The endometrium was proliferative. Pathology for the associated tumor tissue indicated multiple (3 intramural, 1 subserosal) leiomyomata. The patient presented with abdominal pain and infertility. Patient history included scoliosis. Family history included hyperlipidemia, benign hypertension, atherosclerotic coronary artery disease, depressive disorder, brain cancer, and type II diabetes. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS(1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48-hours/round) reannealing hybridization was used.

Table 6 (cont.)

Library	Vector	Library Description
PROSTMT01	pINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 67-year-old Caucasian male during radical prostatectomy with regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated grade 3, Gleason grade 3+3 adenocarcinoma. The patient presented elevated prostate specific antigen (PSA) and induration. Patient history included hyperlipidemia cerebrovascular disease, and a depressive disorder. Family history included atherosclerotic coronary artery disease and hyperlipidemia.
SININOT04	pINCY	The SININOT04 library was constructed using RNA isolated from diseased ileum tissue obtained from a 26-year-old Caucasian male during a partial colectomy, permanent colostomy, and an incidental appendectomy. Pathology indicated moderately to severely active Crohn's disease. Family history included enteritis of the small intestine.
TESTNOT17	pINCY	Library was constructed using 1.5 micrograms of polyA RNA isolated from testis tissue removed from a 26-year-old Caucasian male who died from head trauma due to a motor vehicle accident. Serologies were negative. Patient history included a hernia at birth, tobacco use (1 1/2 ppd), marijuana use, and daily alcohol use (beer and hard liquor). cDNA synthesis was initiated using a NotI-anchored oligo(dT) primer. Double-stranded cDNA was blunted, ligated to EcoRI adaptors, digested with NotI, size-selected, and cloned into the NotI and EcoRI sites of the pINCY vector (Incyte). The library was then linearized and recircularized to select for insert-containing clones as follows: plasmid DNA was prepped from approximately 1 million clones from the testis tissue library following soft agar transformation. The DNA was linearized with NotI and insert-containing clones were size-selected by agarose gel electrophoresis and recircularized by ligation.
BEPINOT01	PSPORT1	Library was constructed using RNA isolated from a bronchial epithelium primary cell line derived from a 54-year-old Caucasian male.
BRAINOT09	pINCY	Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus, who died at 23 weeks' gestation.

Table 6 (cont.)

Library	Vector	Library Description
BRSTTUT14	pINCY	Library was constructed using RNA isolated from breast tumor tissue removed from a 62-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated an invasive grade 3 (of 4), nuclear grade 3 (of 3) adenocarcinoma, ductal type. Ductal carcinoma in situ, comedo type, comprised 60% of the tumor mass. Metastatic adenocarcinoma was identified in one (of 14) axillary lymph nodes with no perinodal extension. Tumor cells were strongly positive for estrogen receptors and weakly positive for progesterone receptors. Patient history included benign colon neoplasm, hyperlipidemia, cardiac dysrhythmia, and obesity. Family history included atherosclerotic coronary artery disease, myocardial infarction, colon cancer, ovarian cancer, lung cancer, and cerebrovascular disease.
CONUTUT01	pINCY	Library was constructed using RNA isolated from sigmoid mesentery tumor tissue obtained from a 61-year-old female during a total abdominal hysterectomy and bilateral salpingo-oophorectomy with regional lymph node excision. Pathology indicated a metastatic grade 4 malignant mixed mullerian tumor present in the sigmoid mesentery at two sites.
THYRNOT09	pINCY	Library was constructed using RNA isolated from diseased thyroid tissue removed from an 18-year-old Caucasian female during an unilateral thyroid lobectomy and regional lymph node excision. Pathology indicated adenomatous goiter. This was associated with a follicular adenoma of the thyroid. Family history included thyroid cancer in the father.
URETTUT01	pINCY	Library was constructed using RNA isolated from right ureter tumor tissue of a 69-year-old Caucasian male during ureterectomy and lymph node excision. Pathology indicated invasive grade 3 transitional cell carcinoma. Patient history included benign colon neoplasm, tobacco use, asthma, emphysema, acute duodenal ulcer, and hyperplasia of the prostate. Family history included atherosclerotic coronary artery disease, congestive heart failure, and malignant lung neoplasm.
BONRFET01	pINCY	Library was constructed using RNA isolated from rib bone tissue removed from a Caucasian male fetus, who died from Patau's syndrome (trisomy 13) at 20-weeks' gestation.

Table 6 (cont.)

Library	Vector	Library Description
BONSTUT01	pINCY	Library was constructed using RNA isolated from sacral bone tumor tissue removed from an 18-year-old Caucasian female during an exploratory laparotomy with soft tissue excision. Pathology indicated giant cell tumor of the sacrum. Patient history included a soft tissue malignant neoplasm. Family history included prostate cancer.
FIBPFEN06	pINCY	The normalized prostate stromal fibroblast tissue libraries were constructed from 1.56 million independent clones from a fibroblast library. Starting RNA was made from fibroblasts of prostate stroma removed from a male fetus, who died after 26 weeks' gestation. The libraries were normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48-hours/round) reannealing hybridization was used. The library was then linearized and recircularized to select for insert containing clones as follows: plasmid DNA was prepped from approximately 1 million clones from the normalized prostate stromal fibroblast tissue libraries following soft agar transformation. The DNA was linearized with NotI and insert containing clones were size-selected by agarose gel electrophoresis and then recircularized by ligation.
KIDNNOT05	PSPORT1	Library was constructed using RNA isolated from the kidney tissue of a 2-day-old Hispanic female, who died from cerebral anoxia. Family history included congenital heart disease.
LUNGTUT09	pINCY	Library was constructed using RNA isolated from lung tumor tissue removed from a 68-year-old Caucasian male during segmental lung resection. Pathology indicated invasive grade 3 squamous cell carcinoma and a metastatic tumor. Patient history included type II diabetes, thyroid disorder, depressive disorder, hyperlipidemia, esophageal ulcer, and tobacco use.
PROSBPT07	pINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 53-year-old Caucasian male during radical prostatectomy and regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated adenocarcinoma (Gleason grade 3+2). The patient presented with elevated prostate specific antigen and induration. Patient history included hyperlipidemia. Family history included atherosclerotic coronary artery disease, coronary artery bypass graft, perforated gallbladder, hyperlipidemia, and kidney stones.

Table 6 (cont.)

Library	Vector	Library Description
SINITME01	pINCY	This 5' biased random primed library was constructed using RNA isolated from ileum tissue removed from a 70-year-old Caucasian female during right hemicolectomy, open liver biopsy, flexible sigmoidoscopy, colonoscopy, and permanent colostomy. Pathology for the matched tumor tissue indicated invasive grade 2 adenocarcinoma forming an ulcerated mass, situated 2 cm distal to the ileocecal valve. The tumor invaded through the muscularis propria just into the serosal adipose tissue. One (of 16) regional lymph node was positive for a microfocus of metastatic adenocarcinoma. Focal fat necrosis was identified from pelvic region tissue. Patient history included a malignant breast neoplasm, type II diabetes, hyperlipidemia, viral hepatitis, an unspecified thyroid disorder, osteoarthritis, a malignant skin neoplasm, deficiency anemia, and normal delivery. Family history included breast cancer, atherosclerotic coronary artery disease, benign hypertension, cerebrovascular disease, ovarian cancer, and hyperlipidemia.
SINTNOR01	PCDNA2.1	This random primed library was constructed using RNA isolated from small intestine tissue removed from a 31-year-old Caucasian female during Roux-en-Y gastric bypass. Patient history included clinical obesity.
THP1AZT01	pINCY	Library was constructed using polyA RNA isolated from THP-1 promonocyte cells treated for three days with 0.8 micromolar 5-aza-2'-deoxycytidine. THP-1 (ATCC TIB 202) is a human promonocyte line derived from peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (ref: Int. J. Cancer (1980) 26:171).

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits:* Probability value= 1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-24,

b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-24,

c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24, and

d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-24.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-24.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:25-48.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:25-48,

b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:25-48,

c) a polynucleotide sequence complementary to a),

d) a polynucleotide sequence complementary to b), and

e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-24.

18. A method for treating a disease or condition associated with decreased expression of functional DME, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional DME, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional DME, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

SEQUENCE LISTING

<110> INCYTE GENOMICS, INC.
YANG, Junming
BAUGHN, Mariah R.
BURFORD, Neil
AU-YOUNG, Janice
LU, Dyung Aina M.
REDDY, Roopa
RING, Huijun Z.
HILLMAN, Jennifer L.
YUE, Henry
AZIMZAI, Yalda
YAO, Monique G.
GANDHI, Ameena R.
NGUYEN, Danniel B.
BAUGHN, Mariah R.
TANG, Y. Tom
LAL, Preeti
YUE, Henry
BANDMAN, Olga

<120> DRUG METABOLIZING ENZYMES

<130> PI-0007 PCT

<140> To Be Assigned
<141> Herewith

<160> 48
<170> PERL Program

<210> 1
<211> 330
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1799250CD1

<400> 1

Met	Ser	Pro	Leu	Ser	Ala	Ala	Arg	Ala	Ala	Leu	Arg	Val	Tyr	Ala	
1				5					10						15
Val	Gly	Ala	Ala	Val	Ile	Leu	Ala	Gln	Leu	Leu	Arg	Arg	Cys	Arg	
				20					25						30
Gly	Gly	Phe	Leu	Glu	Pro	Val	Leu	Pro	Pro	Arg	Pro	Asp	Arg	Val	
				35					40						45
Ala	Ile	Val	Thr	Gly	Gly	Thr	Asp	Gly	Ile	Gly	Tyr	Ser	Thr	Ala	
				50					55						60
Lys	His	Leu	Ala	Arg	Leu	Gly	Met	His	Val	Ile	Ile	Ala	Gly	Asn	
				65					70						75
Asn	Asp	Ser	Lys	Ala	Lys	Gln	Val	Val	Ser	Lys	Ile	Lys	Glu	Glu	
				80					85						90
Thr	Leu	Asn	Asp	Lys	Val	Glu	Phe	Leu	Tyr	Cys	Asp	Leu	Ala	Ser	
				95					100						105
Met	Thr	Ser	Ile	Arg	Gln	Phe	Val	Gln	Lys	Phe	Lys	Met	Lys	Lys	
				110					115						120
Ile	Pro	Leu	His	Val	Leu	Ile	Asn	Asn	Ala	Gly	Val	Met	Met	Val	
				125					130						135
Pro	Gln	Arg	Lys	Thr	Arg	Asp	Gly	Phe	Glu	Glu	His	Phe	Gly	Leu	
				140					145						150
Asn	Tyr	Leu	Gly	His	Phe	Leu	Leu	Thr	Asn	Leu	Leu	Leu	Asp	Thr	

```
                155                 160                 165
Leu Lys Glu Ser Gly Ser Pro Gly His Ser Ala Arg Val Val Thr
                170                 175                 180
Val Ser Ser Ala Thr His Tyr Val Ala Glu Leu Asn Met Asp Asp
                185                 190                 195
Leu Gln Ser Ser Ala Cys Tyr Ser Pro His Ala Ala Tyr Ala Gln
                200                 205                 210
Ser Lys Leu Ala Leu Val Leu Phe Thr Tyr His Leu Gln Arg Leu
                215                 220                 225
Leu Ala Ala Glu Gly Ser His Val Thr Ala Asn Val Val Asp Pro
                230                 235                 240
Gly Val Val Asn Thr Asp Val Tyr Lys His Val Phe Trp Ala Thr
                245                 250                 255
Arg Leu Ala Lys Lys Leu Leu Gly Trp Leu Leu Phe Lys Thr Pro
                260                 265                 270
Asp Glu Gly Ala Trp Thr Ser Ile Tyr Ala Ala Val Thr Pro Glu
                275                 280                 285
Leu Glu Gly Val Gly Gly Arg Tyr Leu Tyr Asn Lys Lys Glu Thr
                290                 295                 300
Lys Ser Leu His Val Thr Tyr Asn Gln Lys Leu Gln Gln Gln Leu
                305                 310                 315
Trp Ser Lys Ser Cys Glu Met Thr Gly Val Leu Asp Val Thr Leu
                320                 325                 330
```

<210> 2
<211> 497
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2242475CD1

<400> 2

```
Met Gly Leu Glu Ala Leu Val Pro Leu Ala Val Ile Val Ala Ile
  1                 5                  10                  15
Phe Leu Leu Leu Val Asp Leu Met His Arg Arg Gln Arg Trp Ala
                 20                  25                  30
Ala Arg Tyr Pro Pro Gly Pro Leu Pro Leu Pro Gly Leu Gly Asn
                 35                  40                  45
Leu Leu His Val Asp Phe Gln Asn Thr Pro Tyr Cys Phe Asp Gln
                 50                  55                  60
Leu Arg Arg Arg Phe Gly Asp Val Phe Ser Leu Gln Leu Ala Trp
                 65                  70                  75
Thr Pro Val Val Val Leu Asn Gly Leu Ala Ala Val Arg Glu Ala
                 80                  85                  90
Leu Val Thr His Gly Glu Asp Thr Ala Asp Arg Pro Pro Val Pro
                 95                 100                 105
Ile Thr Gln Ile Leu Gly Phe Gly Pro Arg Ser Gln Gly Val Phe
                110                 115                 120
Leu Ala Arg Tyr Gly Pro Ala Trp Arg Glu Gln Arg Arg Phe Ser
                125                 130                 135
Val Ser Thr Leu Arg Asn Leu Gly Leu Gly Lys Lys Ser Leu Glu
                140                 145                 150
Gln Trp Val Thr Glu Glu Ala Ala Cys Leu Cys Ala Ala Phe Ala
                155                 160                 165
Asn His Ser Gly Arg Pro Phe Arg Pro Asn Gly Leu Leu Asp Lys
                170                 175                 180
Ala Val Ser Asn Val Ile Ala Ser Leu Thr Cys Gly Arg Arg Phe
                185                 190                 195
Glu Tyr Asp Asp Pro Arg Phe Leu Arg Leu Leu Asp Leu Ala Gln
                200                 205                 210
Glu Gly Leu Lys Glu Glu Ser Gly Phe Leu Arg Glu Val Leu Asn
```

215 220 225
Ala Val Pro Val Leu Pro His Ile Pro Ala Leu Ala Gly Lys Val
230 235 240
Leu Arg Phe Gln Lys Ala Phe Leu Thr Gln Leu Asp Glu Leu Leu
245 250 255
Thr Glu His Arg Met Thr Trp Asp Pro Ala Gln Pro Pro Arg Asp
260 265 270
Leu Thr Glu Ala Phe Leu Ala Lys Lys Glu Lys Ala Lys Gly Ser
275 280 285
Pro Glu Ser Ser Phe Asn Asp Glu Asn Leu Arg Ile Val Val Gly
290 295 300
Asn Leu Phe Leu Ala Gly Met Val Thr Thr Ser Thr Thr Leu Ala
305 310 315
Trp Ala Leu Leu Leu Met Ile Leu His Pro Asp Val Gln Cys Arg
320 325 330
Val Gln Gln Glu Ile Asp Glu Val Ile Gly Gln Val Arg His Pro
335 340 345
Glu Met Ala Asp Gln Ala His Met Pro Phe Thr Asn Ala Val Ile
350 355 360
His Glu Val Gln Arg Phe Ala Asp Ile Val Pro Met Asn Leu Pro
365 370 375
His Lys Thr Ser Arg Asp Ile Glu Val Gln Gly Phe Leu Ile Pro
380 385 390
Lys Gly Thr Thr Leu Ile Pro Asn Leu Ser Ser Val Leu Lys Asp
395 400 405
Glu Thr Val Trp Glu Lys Pro Leu Arg Phe His Pro Glu His Phe
410 415 420
Leu Asp Ala Gln Gly Asn Phe Val Lys His Glu Ala Phe Met Pro
425 430 435
Phe Ser Ala Gly Arg Arg Ala Cys Leu Gly Glu Pro Leu Ala Arg
440 445 450
Met Glu Leu Phe Leu Phe Phe Thr Cys Leu Leu Gln Arg Phe Ser
455 460 465
Phe Ser Val Pro Thr Gly Gln Pro Arg Pro Ser Asp Tyr Gly Val
470 475 480
Phe Ala Phe Leu Leu Ser Pro Ser Pro Tyr Gln Leu Cys Ala Phe
485 490 495
Lys Arg

<210> 3
<211> 286
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2706492CD1

<400> 3
Met Ala Leu Asp Leu Ala Ser Leu Ala Ser Val Arg Ala Phe Ala
1 5 10 15
Thr Ala Phe Leu Ser Ser Glu Pro Arg Leu Asp Ile Leu Ile His
20 25 30
Asn Ala Gly Ile Ser Ser Cys Gly Arg Thr Arg Glu Ala Phe Asn
35 40 45
Leu Leu Leu Arg Val Asn His Ile Gly Pro Phe Leu Leu Thr His
50 55 60
Leu Leu Leu Pro Cys Leu Lys Ala Cys Ala Pro Ser Arg Val Val
65 70 75
Val Val Ala Ser Ala Ala His Cys Arg Gly Arg Leu Asp Phe Lys
80 85 90
Arg Leu Asp Arg Pro Val Val Gly Trp Arg Gln Glu Leu Arg Ala
95 100 105

```
Tyr Ala Asp Thr Lys Leu Ala Asn Val Leu Phe Ala Arg Glu Leu
                110                 115                 120
Ala Asn Gln Leu Glu Ala Thr Gly Val Thr Cys Tyr Ala Ala His
                125                 130                 135
Pro Gly Pro Val Asn Ser Glu Leu Phe Leu Arg His Val Pro Gly
                140                 145                 150
Trp Leu Arg Pro Leu Leu Arg Pro Leu Ala Trp Leu Val Leu Arg
                155                 160                 165
Ala Pro Arg Gly Gly Ala Gln Thr Pro Leu Tyr Cys Ala Leu Gln
                170                 175                 180
Glu Gly Ile Glu Pro Leu Ser Gly Arg Tyr Phe Ala Asn Cys His
                185                 190                 195
Val Glu Glu Val Pro Pro Ala Ala Arg Asp Asp Arg Ala Ala His
                200                 205                 210
Arg Leu Trp Glu Ala Ser Lys Arg Leu Ala Gly Leu Gly Pro Gly
                215                 220                 225
Glu Asp Ala Glu Pro Asp Glu Asp Pro Gln Ser Glu Asp Ser Glu
                230                 235                 240
Ala Pro Ser Ser Leu Ser Thr Pro His Pro Glu Glu Pro Thr Val
                245                 250                 255
Ser Gln Pro Tyr Pro Ser Pro Gln Ser Ser Pro Asp Leu Ser Lys
                260                 265                 270
Met Thr His Arg Ile Gln Ala Lys Val Glu Pro Glu Ile Gln Leu
                275                 280                 285
Ser
```

<210> 4
<211> 469
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2766688CD1

<400> 4

```
Met Glu Leu Ile Ser Pro Thr Val Ile Ile Ile Leu Gly Cys Leu
  1                 5                  10                  15
Ala Leu Phe Leu Leu Leu Gln Arg Lys Asn Leu Arg Arg Pro Pro
                 20                  25                  30
Cys Ile Lys Gly Trp Ile Pro Trp Ile Gly Val Gly Phe Glu Phe
                 35                  40                  45
Gly Lys Ala Pro Leu Glu Phe Ile Glu Lys Ala Arg Ile Lys Tyr
                 50                  55                  60
Gly Pro Ile Phe Thr Val Phe Ala Met Gly Asn Arg Met Thr Phe
                 65                  70                  75
Val Thr Glu Glu Glu Gly Ile Asn Val Phe Leu Lys Ser Lys Lys
                 80                  85                  90
Val Asp Phe Glu Leu Ala Val Gln Asn Ile Val Tyr His Thr Ala
                 95                 100                 105
Ser Ile Pro Lys Asn Val Phe Leu Ala Leu His Glu Lys Leu Tyr
                110                 115                 120
Ile Met Leu Lys Gly Lys Met Gly Thr Val Asn Leu His Gln Phe
                125                 130                 135
Thr Gly Gln Leu Thr Glu Glu Leu His Glu Gln Leu Glu Asn Leu
                140                 145                 150
Gly Thr His Gly Thr Met Asp Leu Asn Asn Leu Val Arg His Leu
                155                 160                 165
Leu Tyr Pro Val Thr Val Asn Met Leu Phe Asn Lys Ser Leu Phe
                170                 175                 180
Ser Thr Asn Lys Lys Lys Ile Lys Glu Phe His Gln Tyr Phe Gln
                185                 190                 195
Val Tyr Asp Glu Asp Phe Glu Tyr Gly Ser Gln Leu Pro Glu Cys
```

```
                200                 205                 210
Leu Leu Arg Asn Trp Ser Lys Ser Lys Lys Trp Phe Leu Glu Leu
                215                 220                 225
Phe Glu Lys Asn Ile Pro Asp Ile Lys Ala Cys Lys Ser Ala Lys
                230                 235                 240
Asp Asn Ser Met Thr Leu Leu Gln Ala Thr Leu Asp Ile Val Glu
                245                 250                 255
Thr Glu Thr Ser Lys Glu Asn Ser Pro Asn Tyr Gly Leu Leu Leu
                260                 265                 270
Leu Trp Ala Ser Leu Ser Asn Ala Val Pro Val Ala Phe Trp Thr
                275                 280                 285
Leu Ala Tyr Val Leu Ser His Pro Asp Ile His Lys Ala Ile Met
                290                 295                 300
Glu Gly Ile Ser Ser Val Phe Gly Lys Ala Gly Lys Asp Lys Ile
                305                 310                 315
Lys Val Ser Glu Asp Asp Leu Glu Asn Leu Leu Leu Ile Lys Trp
                320                 325                 330
Cys Val Leu Glu Thr Ile Arg Leu Lys Ala Pro Gly Val Ile Thr
                335                 340                 345
Arg Lys Val Val Lys Pro Val Glu Ile Leu Asn Tyr Ile Ile Pro
                350                 355                 360
Ser Gly Asp Leu Leu Met Leu Ser Pro Phe Trp Leu His Arg Asn
                365                 370                 375
Pro Lys Tyr Phe Pro Glu Pro Glu Leu Phe Lys Pro Glu Arg Trp
                380                 385                 390
Glu Lys Gly Lys Phe Arg Glu Ala Leu Phe Leu Gly Leu Leu His
                395                 400                 405
Gly Ile Gly Ser Gly Lys Phe Gln Cys Pro Ala Arg Trp Phe Ala
                410                 415                 420
Leu Leu Glu Val Gln Met Cys Ile Ile Leu Ile Leu Tyr Lys Tyr
                425                 430                 435
Asp Cys Ser Leu Leu Asp Pro Leu Pro Lys Gln Ser Tyr Leu His
                440                 445                 450
Leu Val Gly Val Pro Gln Pro Glu Gly Gln Cys Arg Ile Glu Tyr
                455                 460                 465
Lys Gln Arg Ile
```

<210> 5
<211> 331
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2788823CD1

<400> 5

```
Met Ser Arg Tyr Leu Leu Pro Leu Ser Ala Leu Gly Thr Val Ala
  1                 5                  10                  15
Gly Ala Ala Val Leu Leu Lys Asp Tyr Val Thr Gly Gly Ala Cys
                 20                  25                  30
Pro Ser Lys Ala Thr Ile Pro Gly Lys Thr Val Ile Val Thr Gly
                 35                  40                  45
Ala Asn Thr Gly Ile Gly Lys Gln Thr Ala Leu Glu Leu Ala Arg
                 50                  55                  60
Arg Gly Gly Asn Ile Ile Leu Ala Cys Arg Asp Met Glu Lys Cys
                 65                  70                  75
Glu Ala Ala Ala Lys Asp Ile Arg Gly Glu Thr Leu Asn His His
                 80                  85                  90
Val Asn Ala Arg His Leu Asp Leu Ala Ser Leu Lys Ser Ile Arg
                 95                 100                 105
Glu Phe Ala Ala Lys Ile Ile Glu Glu Glu Glu Arg Val Asp Ile
                110                 115                 120
```

```
Leu Ile Asn Asn Ala Gly Val Met Arg Cys Pro His Trp Thr Thr
                125                 130                 135
Glu Asp Gly Phe Glu Met Gln Phe Gly Val Asn His Leu Gly His
                140                 145                 150
Phe Leu Leu Thr Asn Leu Leu Leu Asp Lys Leu Lys Ala Ser Ala
                155                 160                 165
Pro Ser Arg Ile Ile Asn Leu Ser Ser Leu Ala His Val Ala Gly
                170                 175                 180
His Ile Asp Phe Asp Asp Leu Asn Trp Gln Thr Arg Lys Tyr Asn
                185                 190                 195
Thr Lys Ala Ala Tyr Cys Gln Ser Lys Leu Ala Ile Val Leu Phe
                200                 205                 210
Thr Lys Glu Leu Ser Arg Arg Leu Gln Gly Ser Gly Val Thr Val
                215                 220                 225
Asn Ala Leu His Pro Gly Val Ala Arg Thr Glu Leu Gly Arg His
                230                 235                 240
Thr Gly Ile His Gly Ser Thr Phe Ser Ser Thr Thr Leu Gly Pro
                245                 250                 255
Ile Phe Trp Leu Leu Val Lys Ser Pro Glu Leu Ala Ala Gln Pro
                260                 265                 270
Ser Thr Tyr Leu Ala Val Ala Glu Glu Leu Ala Asp Val Ser Gly
                275                 280                 285
Lys Tyr Phe Asp Gly Leu Lys Gln Lys Ala Pro Ala Pro Glu Ala
                290                 295                 300
Glu Asp Glu Glu Val Ala Arg Arg Leu Trp Ala Glu Ser Ala Arg
                305                 310                 315
Leu Val Gly Leu Glu Ala Pro Ser Val Arg Glu Gln Pro Leu Pro
                320                 325                 330
Arg
```

<210> 6
<211> 509
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3348822CD1

<400> 6

```
Met Glu Phe Ser Trp Leu Glu Thr Arg Trp Ala Arg Pro Phe Tyr
  1               5                  10                  15
Leu Ala Phe Val Phe Cys Leu Ala Leu Gly Leu Leu Gln Ala Ile
                20                  25                  30
Lys Leu Tyr Leu Arg Arg Gln Arg Leu Leu Arg Asp Leu Arg Pro
                35                  40                  45
Phe Pro Ala Pro Pro Thr His Trp Phe Leu Gly His Gln Lys Phe
                50                  55                  60
Ile Gln Asp Asp Asn Met Glu Lys Leu Glu Glu Ile Ile Glu Lys
                65                  70                  75
Tyr Pro Arg Ala Phe Pro Phe Trp Ile Gly Pro Phe Gln Ala Phe
                80                  85                  90
Phe Cys Ile Tyr Asp Pro Asp Tyr Ala Lys Thr Leu Leu Ser Arg
                95                 100                 105
Thr Asp Pro Lys Ser Gln Tyr Leu Gln Lys Phe Ser Pro Pro Leu
                110                 115                 120
Leu Gly Lys Gly Leu Ala Ala Leu Asp Gly Pro Lys Trp Phe Gln
                125                 130                 135
His Arg Arg Leu Leu Thr Pro Gly Phe His Phe Asn Ile Leu Lys
                140                 145                 150
Ala Tyr Ile Glu Val Met Ala His Ser Val Lys Met Met Leu Asp
                155                 160                 165
Lys Trp Glu Lys Ile Cys Ser Thr Gln Asp Thr Ser Val Glu Val
```

```
                170                 175                 180
Tyr Glu His Ile Asn Ser Met Ser Leu Asp Ile Ile Met Lys Cys
                185                 190                 195
Ala Phe Ser Lys Glu Thr Asn Cys Gln Thr Asn Ser Thr His Asp
                200                 205                 210
Pro Tyr Ala Lys Ala Ile Phe Glu Leu Ser Lys Ile Ile Phe His
                215                 220                 225
Arg Leu Tyr Ser Leu Leu Tyr His Ser Asp Ile Ile Phe Lys Leu
                230                 235                 240
Ser Pro Gln Gly Tyr Arg Phe Gln Lys Leu Ser Arg Val Leu Asn
                245                 250                 255
Gln Tyr Thr Asp Thr Ile Ile Gln Glu Arg Lys Lys Ser Leu Gln
                260                 265                 270
Ala Gly Val Lys Gln Asp Asn Thr Pro Lys Arg Lys Tyr Gln Asp
                275                 280                 285
Phe Leu Asp Ile Val Leu Ser Ala Lys Asp Glu Ser Gly Ser Ser
                290                 295                 300
Phe Ser Asp Ile Asp Val His Ser Glu Val Ser Thr Phe Leu Leu
                305                 310                 315
Ala Gly His Asp Thr Leu Ala Ala Ser Ile Ser Trp Ile Leu Tyr
                320                 325                 330
Cys Leu Ala Leu Asn Pro Glu His Gln Glu Arg Cys Arg Glu Glu
                335                 340                 345
Val Arg Gly Ile Leu Gly Asp Gly Ser Ser Ile Thr Trp Asp Gln
                350                 355                 360
Leu Gly Glu Met Ser Tyr Thr Thr Met Cys Ile Lys Glu Thr Cys
                365                 370                 375
Arg Leu Ile Pro Ala Val Pro Ser Ile Ser Arg Asp Leu Ser Lys
                380                 385                 390
Pro Leu Thr Phe Pro Asp Gly Cys Thr Leu Pro Ala Gly Ile Thr
                395                 400                 405
Val Val Leu Ser Ile Trp Gly Leu His His Asn Pro Ala Val Trp
                410                 415                 420
Lys Asn Pro Lys Val Phe Asp Pro Leu Arg Phe Ser Gln Glu Asn
                425                 430                 435
Ser Asp Gln Arg His Pro Tyr Ala Tyr Leu Pro Phe Ser Ala Gly
                440                 445                 450
Ser Arg Asn Cys Ile Gly Gln Glu Phe Ala Met Ile Glu Leu Lys
                455                 460                 465
Val Thr Ile Ala Leu Ile Leu Leu His Phe Arg Val Thr Pro Asp
                470                 475                 480
Pro Thr Arg Pro Leu Thr Phe Pro Asn His Phe Ile Leu Lys Pro
                485                 490                 495
Lys Asn Gly Met Tyr Leu His Leu Lys Lys Leu Ser Glu Cys
                500                 505
```

<210> 7
<211> 433
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4290251CD1

<400> 7

```
Met Ala Gly Thr Asn Ala Leu Leu Met Leu Glu Asn Phe Ile Asp
  1               5                  10                  15
Gly Lys Phe Leu Pro Cys Ser Ser Tyr Ile Asp Ser Tyr Asp Pro
                20                  25                  30
Ser Thr Gly Glu Val Tyr Cys Arg Val Pro Asn Ser Gly Lys Asp
                35                  40                  45
Glu Ile Glu Ala Ala Val Lys Ala Ala Arg Glu Ala Phe Pro Ser
                50                  55                  60
```

```
Trp Ser Ser Arg Ser Pro Gln Glu Arg Ser Arg Val Leu Asn Gln
                 65                  70                  75
Val Ala Asp Leu Leu Glu Gln Ser Leu Glu Glu Phe Ala Gln Ala
                 80                  85                  90
Glu Ser Lys Asp Gln Gly Lys Thr Leu Ala Leu Ala Arg Thr Met
                 95                 100                 105
Asp Ile Pro Arg Ser Val Gln Asn Phe Arg Phe Phe Ala Ser Ser
                110                 115                 120
Ser Leu His His Thr Ser Glu Cys Thr Gln Met Glu His Leu Gly
                125                 130                 135
Cys Met His Tyr Thr Val Arg Ala Pro Val Gly Val Ala Gly Leu
                140                 145                 150
Ile Ser Pro Trp Asn Leu Pro Leu Tyr Leu Leu Thr Trp Lys Ile
                155                 160                 165
Ala Pro Ala Met Ala Ala Gly Asn Thr Val Ile Ala Lys Pro Ser
                170                 175                 180
Glu Leu Thr Ser Val Thr Ala Trp Met Leu Cys Lys Leu Leu Asp
                185                 190                 195
Lys Ala Gly Val Pro Pro Gly Val Val Asn Ile Val Phe Gly Thr
                200                 205                 210
Gly Pro Arg Val Gly Glu Ala Leu Val Ser His Pro Glu Val Pro
                215                 220                 225
Leu Ile Ser Phe Thr Gly Ser Gln Pro Thr Ala Glu Arg Ile Thr
                230                 235                 240
Gln Leu Ser Ala Pro His Cys Lys Lys Leu Ser Leu Glu Leu Gly
                245                 250                 255
Gly Lys Asn Pro Ala Ile Ile Phe Glu Asp Ala Asn Leu Asp Glu
                260                 265                 270
Cys Ile Pro Ala Thr Val Arg Ser Ser Phe Ala Asn Gln Val Arg
                275                 280                 285
Ser Tyr Val Lys Arg Ala Leu Ala Glu Ser Ala Gln Ile Trp Cys
                290                 295                 300
Gly Glu Gly Val Asp Lys Leu Ser Leu Pro Ala Arg Asn Gln Ala
                305                 310                 315
Gly Tyr Phe Met Leu Pro Thr Val Ile Thr Asp Ile Lys Asp Glu
                320                 325                 330
Ser Cys Cys Met Thr Glu Glu Ile Phe Gly Pro Val Thr Cys Val
                335                 340                 345
Val Pro Phe Asp Ser Glu Glu Glu Val Ile Glu Arg Ala Asn Asn
                350                 355                 360
Val Lys Tyr Gly Leu Ala Ala Thr Val Trp Ser Ser Asn Val Gly
                365                 370                 375
Arg Val His Arg Val Ala Lys Lys Leu Gln Ser Gly Leu Val Trp
                380                 385                 390
Thr Asn Cys Trp Leu Ile Arg Glu Leu Asn Leu Pro Phe Gly Gly
                395                 400                 405
Met Lys Ser Ser Gly Ile Gly Arg Glu Gly Ala Lys Asp Ser Tyr
                410                 415                 420
Asp Phe Phe Thr Glu Ile Lys Thr Ile Thr Val Lys His
                425                 430
```

<210> 8
<211> 186
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4904188CD1

<400> 8

```
Met Lys Thr Glu Asp Gly Phe Glu Met Gln Phe Gly Val Asn His
  1               5                  10                  15
Leu Gly His Phe Leu Leu Thr Asn Leu Leu Leu Gly Leu Leu Lys
```

```
                 20                  25                  30
Ser Ser Ala Pro Ser Arg Ile Val Val Val Ser Ser Lys Leu Tyr
                 35                  40                  45
Lys Tyr Gly Asp Ile Asn Phe Asp Asp Leu Asn Ser Glu Gln Ser
                 50                  55                  60
Tyr Asn Lys Ser Phe Cys Tyr Ser Arg Ser Lys Leu Ala Asn Ile
                 65                  70                  75
Leu Phe Thr Arg Glu Leu Ala Arg Arg Leu Glu Gly Thr Asn Val
                 80                  85                  90
Thr Val Asn Val Leu His Pro Gly Ile Val Arg Thr Asn Leu Gly
                 95                 100                 105
Arg His Ile His Ile Pro Leu Leu Val Lys Pro Leu Phe Asn Leu
                110                 115                 120
Val Ser Trp Ala Phe Phe Lys Thr Pro Val Glu Gly Ala Gln Thr
                125                 130                 135
Ser Ile Tyr Leu Ala Ser Ser Pro Glu Val Glu Gly Val Ser Gly
                140                 145                 150
Arg Tyr Phe Gly Asp Cys Lys Glu Glu Glu Leu Leu Pro Lys Ala
                155                 160                 165
Met Asp Glu Ser Val Ala Arg Lys Leu Trp Asp Ile Ser Glu Val
                170                 175                 180
Met Val Gly Leu Leu Lys
                185

<210> 9
<211> 304
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 638419CD1

<400> 9
Met Ala Lys Ile Glu Lys Asn Ala Pro Thr Met Glu Lys Lys Pro
  1                  5                  10                  15
Glu Leu Phe Asn Ile Met Glu Val Asp Gly Val Pro Thr Leu Ile
                 20                  25                  30
Leu Ser Lys Glu Trp Trp Glu Lys Val Cys Asn Phe Gln Ala Lys
                 35                  40                  45
Pro Asp Asp Leu Ile Leu Ala Thr Tyr Pro Lys Ser Gly Thr Thr
                 50                  55                  60
Trp Met His Glu Ile Leu Asp Met Ile Leu Asn Asp Gly Asp Val
                 65                  70                  75
Glu Lys Cys Lys Arg Ala Gln Thr Leu Asp Arg His Ala Phe Leu
                 80                  85                  90
Glu Leu Lys Phe Pro His Lys Glu Lys Pro Asp Leu Glu Phe Val
                 95                 100                 105
Leu Glu Met Ser Ser Pro Gln Leu Ile Lys Thr His Leu Pro Ser
                110                 115                 120
His Leu Ile Pro Pro Ser Ile Trp Lys Glu Asn Cys Lys Ile Val
                125                 130                 135
Tyr Val Ala Arg Asn Pro Lys Asp Cys Leu Val Ser Tyr Tyr His
                140                 145                 150
Phe His Arg Met Ala Ser Phe Met Pro Asp Pro Gln Asn Leu Glu
                155                 160                 165
Glu Phe Tyr Glu Lys Phe Met Ser Gly Lys Val Val Gly Arg Ser
                170                 175                 180
Trp Phe Asp His Val Lys Gly Trp Trp Ala Ala Lys Asp Thr His
                185                 190                 195
Arg Ile Leu Tyr Leu Phe Tyr Glu Asp Ile Lys Lys Asn Pro Lys
                200                 205                 210
His Glu Ile His Lys Val Leu Glu Phe Leu Glu Lys Thr Leu Ser
                215                 220                 225
```

```
Gly Asp Val Ile Asn Lys Ile Val His His Thr Ser Phe Asp Val
                230                 235                 240
Met Lys Asp Asn Pro Met Ala Asn His Thr Ala Val Pro Ala His
                245                 250                 255
Ile Phe Asn His Ser Ile Ser Lys Phe Met Arg Lys Gly Met Pro
                260                 265                 270
Gly Asp Trp Lys Asn His Phe Thr Val Ala Met Asn Glu Asn Phe
                275                 280                 285
Asp Lys His Tyr Glu Lys Lys Met Ala Gly Ser Thr Leu Asn Phe
                290                 295                 300
Cys Leu Glu Ile
```

<210> 10
<211> 629
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1844394CD1

<400> 10

```
Met Lys Leu Gln Asn Leu Phe Val Asp Asp Ser Gly Arg Tyr Leu
  1               5                  10                  15
Ala Ile Gln Phe His Leu Glu Cys Ala Tyr Val Phe Leu Tyr Tyr
                20                  25                  30
Tyr Glu Tyr Arg Lys Ala Lys Asp Gln Leu Asp Ile Ala Lys Asp
                35                  40                  45
Ile Ser Gln Leu Gln Ile Asp Leu Thr Gly Ala Leu Gly Lys Arg
                50                  55                  60
Thr Arg Phe Gln Glu Asn Tyr Val Ala Gln Leu Ile Leu Asp Val
                65                  70                  75
Arg Arg Glu Gly Asp Val Leu Ser Asn Cys Glu Phe Thr Pro Ala
                80                  85                  90
Pro Thr Pro Gln Glu His Leu Thr Lys Asn Leu Glu Leu Asn Asp
                95                 100                 105
Asp Thr Ile Leu Asn Asp Ile Lys Leu Ala Asp Cys Glu Gln Phe
               110                 115                 120
Gln Met Pro Asp Leu Cys Ala Glu Glu Ile Ala Ile Ile Leu Gly
               125                 130                 135
Ile Cys Thr Asn Phe Gln Lys Asn Asn Pro Val His Thr Leu Thr
               140                 145                 150
Glu Val Glu Leu Leu Ala Phe Thr Ser Cys Leu Leu Ser Gln Pro
               155                 160                 165
Lys Phe Trp Ala Ile Gln Thr Ser Ala Leu Ile Leu Arg Thr Lys
               170                 175                 180
Leu Glu Lys Gly Ser Thr Arg Arg Val Glu Arg Ala Met Arg Gln
               185                 190                 195
Thr Gln Ala Leu Ala Asp Gln Phe Glu Asp Lys Thr Thr Ser Val
               200                 205                 210
Leu Glu Arg Leu Lys Ile Phe Tyr Cys Cys Gln Val Pro Pro His
               215                 220                 225
Trp Ala Ile Gln Arg Gln Leu Ala Ser Leu Leu Phe Glu Leu Gly
               230                 235                 240
Cys Thr Ser Ser Ala Leu Gln Ile Phe Glu Lys Leu Glu Met Trp
               245                 250                 255
Glu Asp Val Val Ile Cys Tyr Glu Arg Ala Gly Gln His Gly Lys
               260                 265                 270
Ala Glu Glu Ile Leu Arg Gln Glu Leu Glu Lys Lys Glu Thr Pro
               275                 280                 285
Ser Leu Tyr Cys Leu Leu Gly Asp Val Leu Gly Asp His Ser Cys
               290                 295                 300
Tyr Asp Lys Ala Trp Glu Leu Ser Arg Tyr Arg Ser Ala Arg Ala
```

```
                305                 310                 315
Gln Arg Ser Lys Ala Leu Leu His Leu Arg Asn Lys Glu Phe Gln
                320                 325                 330
Glu Cys Val Glu Cys Phe Glu Arg Ser Val Lys Ile Asn Pro Met
                335                 340                 345
Gln Leu Gly Val Trp Phe Ser Leu Gly Cys Ala Tyr Leu Ala Leu
                350                 355                 360
Glu Asp Tyr Gln Gly Ser Ala Lys Ala Phe Gln Arg Cys Val Thr
                365                 370                 375
Leu Glu Pro Asp Asn Ala Glu Ala Trp Asn Asn Leu Ser Thr Ser
                380                 385                 390
Tyr Ile Arg Leu Lys Gln Lys Val Lys Ala Phe Arg Thr Leu Gln
                395                 400                 405
Glu Ala Leu Lys Cys Asn Tyr Glu His Trp Gln Ile Trp Lys Asn
                410                 415                 420
Tyr Ile Leu Thr Ser Thr Asp Val Gly Glu Phe Ser Glu Ala Ile
                425                 430                 435
Lys Ala Tyr His Arg Leu Leu Asp Leu Arg Asp Lys Tyr Lys Asp
                440                 445                 450
Val Gln Val Leu Lys Ile Leu Val Arg Ala Val Ile Asp Gly Met
                455                 460                 465
Thr Asp Arg Ser Gly Asp Val Ala Thr Gly Leu Lys Gly Lys Leu
                470                 475                 480
Gln Glu Leu Phe Gly Arg Val Thr Ser Arg Val Thr Asn Asp Gly
                485                 490                 495
Glu Ile Trp Arg Leu Tyr Ala His Val Tyr Gly Asn Gly Gln Ser
                500                 505                 510
Glu Lys Pro Asp Glu Asn Glu Lys Ala Phe Gln Cys Leu Ser Lys
                515                 520                 525
Ala Tyr Lys Cys Asp Thr Gln Ser Asn Cys Trp Glu Lys Asp Ile
                530                 535                 540
Thr Ser Phe Lys Glu Val Val Gln Arg Ala Leu Gly Leu Ala His
                545                 550                 555
Val Ala Ile Lys Cys Ser Lys Asn Lys Ser Ser Ser Gln Glu Ala
                560                 565                 570
Val Gln Met Leu Ser Ser Val Arg Leu Asn Leu Arg Gly Leu Leu
                575                 580                 585
Ser Lys Ala Lys Gln Leu Phe Thr Asp Val Ala Thr Gly Glu Met
                590                 595                 600
Ser Arg Glu Leu Ala Asp Asp Ile Thr Ala Met Asp Thr Leu Val
                605                 610                 615
Thr Glu Leu Gln Asp Leu Ser Asn Gln Phe Arg Asn Gln Tyr
                620                 625

<210> 11
<211> 320
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2613056CD1

<400> 11
Met Thr Leu Asp Ser Ile Met Lys Cys Ala Phe Ser His Gln Gly
  1               5                  10                  15
Ser Ile Gln Leu Asp Ser Thr Leu Asp Ser Tyr Leu Lys Ala Val
                20                  25                  30
Phe Asn Leu Ser Lys Ile Ser Asn Gln Arg Met Asn Asn Phe Leu
                35                  40                  45
His His Asn Asp Leu Val Phe Lys Phe Ser Ser Gln Gly Gln Ile
                50                  55                  60
Phe Ser Lys Phe Asn Gln Glu Leu His Gln Phe Thr Glu Lys Val
                65                  70                  75
```

```
Ile Gln Asp Arg Lys Glu Ser Leu Lys Asp Lys Leu Lys Gln Asp
                 80                  85                  90
Thr Thr Gln Lys Arg Arg Trp Asp Phe Leu Asp Ile Leu Leu Ser
                 95                 100                 105
Ala Lys Ser Glu Asn Thr Lys Asp Phe Ser Glu Ala Asp Leu Gln
                110                 115                 120
Ala Glu Val Lys Thr Phe Met Phe Ala Gly His Asp Thr Thr Ser
                125                 130                 135
Ser Ala Ile Ser Trp Ile Leu Tyr Cys Leu Ala Lys Tyr Pro Glu
                140                 145                 150
His Gln Gln Arg Cys Arg Asp Glu Ile Arg Glu Leu Leu Gly Asp
                155                 160                 165
Gly Ser Ser Ile Thr Trp Glu His Leu Ser Gln Met Pro Tyr Thr
                170                 175                 180
Thr Met Cys Ile Lys Glu Cys Leu Arg Leu Tyr Ala Pro Val Val
                185                 190                 195
Asn Ile Ser Arg Leu Leu Asp Lys Pro Ile Thr Phe Pro Asp Gly
                200                 205                 210
Arg Ser Leu Pro Ala Gly Ile Thr Val Phe Ile Asn Ile Trp Ala
                215                 220                 225
Leu His His Asn Pro Tyr Phe Trp Glu Asp Pro Gln Val Phe Asn
                230                 235                 240
Pro Leu Arg Phe Ser Arg Glu Asn Ser Glu Lys Ile His Pro Tyr
                245                 250                 255
Ala Phe Ile Pro Phe Ser Ala Gly Leu Arg Asn Cys Ile Gly Gln
                260                 265                 270
His Phe Ala Ile Ile Glu Cys Lys Val Ala Val Ala Leu Thr Leu
                275                 280                 285
Leu Arg Phe Lys Leu Ala Pro Asp His Ser Arg Pro Pro Gln Pro
                290                 295                 300
Val Arg Gln Val Val Leu Lys Ser Lys Asn Gly Ile His Val Phe
                305                 310                 315
Ala Lys Lys Val Cys
                320

<210> 12
<211> 56
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5053617CD1

<400> 12
Met Ser Gly Cys Pro Asn Cys Val Trp Val Glu Tyr Ala Asp Arg
  1               5                  10                  15
Leu Leu Gln His Phe Gln Asp Gly Gly Glu Arg Ala Leu Ala Ala
                 20                  25                  30
Leu Glu Glu His Val Ala Asp Glu Asn Leu Lys Ala Phe Leu Arg
                 35                  40                  45
Met Glu Ile Arg Leu His Thr Arg Cys Gly Gly
                 50                  55

<210> 13
<211> 377
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5483256CD1

<400> 13
```

```
Met Asp Pro Ala Ala Arg Val Val Arg Ala Leu Trp Pro Gly Gly
 1               5                  10                  15
Cys Ala Leu Ala Trp Arg Leu Gly Gly Arg Pro Gln Pro Leu Leu
            20                  25                  30
Pro Thr Gln Ser Arg Ala Gly Phe Ala Gly Ala Ala Gly Gly Pro
        35                  40                  45
Ser Pro Val Ala Ala Ala Arg Lys Gly Ser Pro Arg Leu Leu Gly
    50                  55                  60
Ala Ala Ala Leu Ala Leu Gly Gly Ala Leu Gly Leu Tyr His Thr
65                  70                  75
Ala Arg Trp His Leu Arg Ala Gln Asp Leu His Ala Glu Arg Ser
                80                  85                  90
Ala Ala Gln Leu Ser Leu Ser Ser Arg Leu Gln Leu Thr Leu Tyr
            95                  100                 105
Gln Tyr Lys Thr Cys Pro Phe Cys Ser Lys Val Arg Ala Phe Leu
        110                 115                 120
Asp Phe His Ala Leu Pro Tyr Gln Val Val Glu Val Asn Pro Val
    125                 130                 135
Arg Arg Ala Glu Ile Lys Phe Ser Ser Tyr Arg Lys Val Pro Ile
140                 145                 150
Leu Val Ala Gln Glu Gly Glu Ser Ser Gln Gln Leu Asn Asp Ser
155                 160                 165
Ser Val Ile Ile Ser Ala Leu Lys Thr Tyr Leu Val Ser Gly Gln
                170                 175                 180
Pro Leu Glu Glu Ile Ile Thr Tyr Tyr Pro Ala Met Lys Ala Val
            185                 190                 195
Asn Glu Gln Gly Lys Glu Val Thr Glu Phe Gly Asn Lys Tyr Trp
        200                 205                 210
Leu Met Leu Asn Glu Lys Glu Ala Gln Gln Val Tyr Gly Gly Lys
    215                 220                 225
Glu Ala Arg Thr Glu Glu Met Lys Trp Arg Gln Trp Ala Asp Asp
230                 235                 240
Trp Leu Val His Leu Ile Ser Pro Asn Val Tyr Arg Thr Pro Thr
                245                 250                 255
Glu Ala Leu Ala Ser Phe Asp Tyr Ile Val Arg Glu Gly Lys Phe
            260                 265                 270
Gly Ala Val Glu Gly Ala Val Ala Lys Tyr Met Gly Ala Ala Ala
        275                 280                 285
Met Tyr Leu Ile Ser Lys Arg Leu Lys Ser Arg His Arg Leu Gln
    290                 295                 300
Asp Asn Val Arg Glu Asp Leu Tyr Glu Ala Ala Asp Lys Trp Val
305                 310                 315
Ala Ala Val Gly Lys Asp Arg Pro Phe Met Gly Gly Gln Lys Pro
                320                 325                 330
Asn Leu Ala Asp Leu Ala Val Tyr Gly Val Leu Arg Val Met Glu
            335                 340                 345
Gly Leu Asp Ala Phe Asp Asp Leu Met Gln His Thr His Ile Gln
        350                 355                 360
Pro Trp Tyr Leu Arg Val Glu Arg Ala Ile Thr Glu Ala Ser Pro
    365                 370                 375
Ala His
```

<210> 14
<211> 501
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5741354CD1

<400> 14
Met Trp Lys Leu Trp Arg Ala Glu Glu Gly Ala Ala Ala Leu Gly

```
  1               5                  10                  15
Gly Ala Leu Phe Leu Leu Leu Phe Ala Leu Gly Val Arg Gln Leu
             20                  25                  30
Leu Lys Gln Arg Arg Pro Met Gly Phe Pro Pro Gly Pro Pro Gly
             35                  40                  45
Leu Pro Phe Ile Gly Asn Ile Tyr Ser Leu Ala Ala Ser Ser Glu
             50                  55                  60
Leu Pro His Val Tyr Met Arg Lys Gln Ser Gln Val Tyr Gly Glu
             65                  70                  75
Ile Phe Ser Leu Asp Leu Gly Gly Ile Ser Thr Val Val Leu Asn
             80                  85                  90
Gly Tyr Asp Val Val Lys Glu Cys Leu Val His Gln Ser Glu Ile
             95                 100                 105
Phe Ala Asp Arg Pro Cys Leu Pro Leu Phe Met Lys Met Thr Lys
            110                 115                 120
Met Gly Gly Leu Leu Asn Ser Arg Tyr Gly Arg Gly Trp Val Asp
            125                 130                 135
His Arg Arg Leu Ala Val Asn Ser Phe Arg Tyr Phe Gly Tyr Gly
            140                 145                 150
Gln Lys Ser Phe Glu Ser Lys Ile Leu Glu Glu Thr Lys Phe Phe
            155                 160                 165
Asn Asp Ala Ile Glu Thr Tyr Lys Gly Arg Pro Phe Asp Phe Lys
            170                 175                 180
Gln Leu Ile Thr Asn Ala Val Ser Asn Ile Thr Asn Leu Ile Ile
            185                 190                 195
Phe Gly Glu Arg Phe Thr Tyr Glu Asp Thr Asp Phe Gln His Met
            200                 205                 210
Ile Glu Leu Phe Ser Glu Asn Val Glu Leu Ala Ala Ser Ala Ser
            215                 220                 225
Val Phe Leu Tyr Asn Ala Phe Pro Trp Ile Gly Ile Leu Pro Phe
            230                 235                 240
Gly Lys His Gln Gln Leu Phe Arg Asn Ala Ala Val Val Tyr Asp
            245                 250                 255
Phe Leu Ser Arg Leu Ile Glu Lys Ala Ser Val Asn Arg Lys Pro
            260                 265                 270
Gln Leu Pro Gln His Phe Val Asp Ala Tyr Leu Asp Glu Met Asp
            275                 280                 285
Gln Gly Lys Asn Asp Pro Ser Ser Thr Phe Ser Lys Glu Asn Leu
            290                 295                 300
Ile Phe Ser Val Gly Glu Leu Ile Ile Ala Gly Thr Glu Thr Thr
            305                 310                 315
Thr Asn Val Leu Arg Trp Ala Ile Leu Phe Met Ala Leu Tyr Pro
            320                 325                 330
Asn Ile Gln Gly Gln Val Gln Lys Glu Ile Asp Leu Ile Met Gly
            335                 340                 345
Pro Asn Gly Lys Pro Ser Trp Asp Asp Lys Cys Lys Met Pro Tyr
            350                 355                 360
Thr Glu Ala Val Leu His Glu Val Leu Arg Phe Cys Asn Ile Val
            365                 370                 375
Pro Leu Gly Ile Phe His Ala Thr Ser Glu Asp Ala Val Val Arg
            380                 385                 390
Gly Tyr Ser Ile Pro Lys Gly Thr Thr Val Ile Thr Asn Leu Tyr
            395                 400                 405
Ser Val His Phe Asp Glu Lys Tyr Trp Arg Asp Pro Glu Val Phe
            410                 415                 420
His Pro Glu Arg Phe Leu Asp Ser Ser Gly Tyr Phe Ala Lys Lys
            425                 430                 435
Glu Ala Leu Val Pro Phe Ser Leu Gly Arg Arg His Cys Leu Gly
            440                 445                 450
Glu His Leu Ala Arg Met Glu Met Phe Leu Phe Phe Thr Ala Leu
            455                 460                 465
Leu Gln Arg Phe His Leu His Phe Pro His Glu Leu Val Pro Asp
            470                 475                 480
Leu Lys Pro Arg Leu Gly Met Thr Leu Gln Pro Gln Pro Tyr Leu
```

485 490 495
Ile Cys Ala Glu Arg Arg
500

<210> 15
<211> 144
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5872615CD1

<400> 15
Met Arg Lys Ile Asp Leu Cys Leu Ser Ser Glu Gly Ser Glu Val
1 5 10 15
Ile Leu Ala Thr Ser Ser Asp Glu Lys His Pro Pro Glu Asn Ile
20 25 30
Ile Asp Gly Asn Pro Glu Thr Phe Trp Thr Thr Thr Gly Met Phe
35 40 45
Pro Gln Glu Phe Ile Ile Cys Phe His Lys His Val Arg Ile Glu
50 55 60
Arg Leu Val Ile Gln Ser Tyr Phe Val Gln Thr Leu Lys Ile Glu
65 70 75
Lys Ser Thr Ser Lys Glu Pro Val Asp Phe Glu Gln Trp Ile Glu
80 85 90
Lys Asp Leu Val His Thr Glu Gly Gln Leu Gln Asn Glu Glu Ile
95 100 105
Val Ala His Asp Gly Ser Ala Thr Tyr Leu Arg Phe Ile Ile Val
110 115 120
Ser Ala Phe Asp His Phe Ala Ser Val His Ser Val Ser Ala Glu
125 130 135
Gly Thr Val Val Ser Asn Leu Ser Ser
140

<210> 16
<211> 218
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2657543CD1

<400> 16
Met Leu Ser Thr Phe Ala Arg Gln Asn Asp Ile Pro Phe Gln Leu
1 5 10 15
Gln Thr Val Glu Leu Ala Trp Gly Glu His Leu Lys Pro Glu Phe
20 25 30
Leu Lys Val Asn Pro Leu Gly Lys Val Pro Ala Leu Arg Asp Gly
35 40 45
Asp Phe Leu Leu Ala Glu Arg Leu Glu Lys Arg Ser Leu Thr Pro
50 55 60
Pro Ala His Ser Met Val Ile Val Leu Tyr Leu Ser Arg Lys Tyr
65 70 75
Gln Ile Arg Gly His Trp Tyr Pro Pro Glu Leu Gln Ala Arg Thr
80 85 90
Cys Val Asp Glu Tyr Leu Ala Trp Lys His Val Thr Ile Gln Leu
95 100 105
Pro Ala Thr Asn Val Tyr Leu Cys Lys Pro Ala Asp Ala Ala Gln
110 115 120
Leu Glu Arg Leu Leu Gly Arg Leu Thr Pro Ala Leu Gln His Leu
125 130 135
Asp Gly Gly Val Leu Val Ala Arg Pro Phe Leu Ala Met Glu Gln

```
                140                 145                 150
Ile Ser Leu Glu Asp Leu Val Leu Thr Glu Val Met Gln Val Lys
                155                 160                 165
Leu Ser Tyr Pro Pro Ala Leu Gly Gly Thr Leu Gly Met Gly Leu
                170                 175                 180
Ser Pro Asn Pro Ser Cys Pro Val Phe Pro Ala His Cys Arg Trp
                185                 190                 195
Leu Arg Pro Leu Pro Arg Leu Ala Leu Ala Gly Ser Val Thr Gly
                200                 205                 210
Pro Tyr Glu Gly Cys Pro Trp Tyr
                215
```

<210> 17
<211> 210
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3041639CD1

<400> 17

```
Met Ala Cys Ile Leu Lys Arg Lys Ser Val Ile Ala Val Ser Phe
  1                 5                  10                  15
Ile Ala Ala Phe Leu Phe Leu Leu Val Val Arg Leu Val Asn Glu
                 20                  25                  30
Val Asn Phe Pro Leu Leu Leu Asn Cys Phe Gly Gln Pro Gly Thr
                 35                  40                  45
Lys Trp Ile Pro Phe Ser Tyr Thr Tyr Arg Arg Pro Leu Arg Thr
                 50                  55                  60
His Tyr Gly Tyr Ile Asn Val Lys Thr Gln Glu Pro Leu Gln Leu
                 65                  70                  75
Asp Cys Asp Leu Cys Ala Ile Val Ser Asn Ser Gly Gln Met Val
                 80                  85                  90
Gly Gln Lys Val Gly Asn Glu Ile Asp Arg Ser Ser Cys Ile Trp
                 95                 100                 105
Arg Met Asn Asn Ala Pro Thr Lys Gly Tyr Glu Glu Asp Val Gly
                110                 115                 120
Arg Met Thr Met Ile Arg Val Val Ser His Thr Ser Val Pro Leu
                125                 130                 135
Leu Leu Lys Asn Pro Asp Tyr Phe Phe Lys Glu Ala Asn Thr Thr
                140                 145                 150
Ile Tyr Val Ile Trp Gly Pro Phe Arg Asn Met Arg Lys Asp Gly
                155                 160                 165
Asn Gly Ile Val Tyr Asn Met Leu Lys Lys Thr Val Gly Ile Tyr
                170                 175                 180
Pro Asn Ala Gln Ile Tyr Val Thr Thr Glu Lys Arg Met Ser Tyr
                185                 190                 195
Cys Asp Gly Val Phe Lys Lys Glu Thr Gly Lys Asp Ser Thr Glu
                200                 205                 210
```

<210> 18
<211> 613
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3595451CD1

<400> 18

```
Met Cys Cys Trp Pro Leu Leu Leu Leu Trp Gly Leu Leu Pro Gly
  1                 5                  10                  15
```

```
Thr Ala Ala Gly Gly Ser Gly Arg Thr Tyr Pro His Arg Thr Leu
                 20                  25                  30
Leu Asp Ser Glu Gly Lys Tyr Trp Leu Gly Trp Ser Gln Arg Gly
                 35                  40                  45
Ser Gln Ile Ala Phe Arg Leu Gln Val Arg Thr Ala Gly Tyr Val
                 50                  55                  60
Gly Phe Gly Phe Ser Pro Thr Gly Ala Met Ala Ser Ala Asp Ile
                 65                  70                  75
Val Val Gly Gly Val Ala His Gly Arg Pro Tyr Leu Gln Asp Tyr
                 80                  85                  90
Phe Thr Asn Ala Asn Arg Glu Leu Lys Lys Asp Ala Gln Gln Asp
                 95                 100                 105
Tyr His Leu Glu Tyr Ala Met Glu Asn Ser Thr His Thr Ile Ile
                110                 115                 120
Glu Phe Thr Arg Glu Leu His Thr Cys Asp Ile Asn Asp Lys Ser
                125                 130                 135
Ile Thr Asp Ser Thr Val Arg Val Ile Trp Ala Tyr His His Glu
                140                 145                 150
Asp Ala Gly Glu Ala Gly Pro Lys Tyr His Asp Ser Asn Arg Gly
                155                 160                 165
Thr Lys Ser Leu Arg Leu Leu Asn Pro Glu Lys Thr Ser Val Leu
                170                 175                 180
Ser Thr Ala Leu Pro Tyr Phe Asp Leu Val Asn Gln Asp Val Pro
                185                 190                 195
Ile Pro Asn Lys Asp Thr Thr Tyr Trp Cys Gln Met Phe Lys Ile
                200                 205                 210
Pro Val Phe Gln Glu Lys His His Val Ile Lys Val Glu Pro Val
                215                 220                 225
Ile Gln Arg Gly His Glu Ser Leu Val His His Ile Leu Leu Tyr
                230                 235                 240
Gln Cys Ser Asn Asn Phe Asn Asp Ser Val Leu Glu Ser Gly His
                245                 250                 255
Glu Cys Tyr His Pro Asn Met Pro Asp Ala Phe Leu Thr Cys Glu
                260                 265                 270
Thr Val Ile Phe Ala Trp Ala Ile Gly Gly Glu Gly Phe Ser Tyr
                275                 280                 285
Pro Pro His Val Gly Leu Ser Leu Gly Thr Pro Leu Asp Pro His
                290                 295                 300
Tyr Val Leu Leu Glu Val His Tyr Asp Asn Pro Thr Tyr Glu Glu
                305                 310                 315
Gly Leu Ile Asp Asn Ser Gly Leu Arg Leu Phe Tyr Thr Met Asp
                320                 325                 330
Ile Arg Lys Tyr Asp Ala Gly Val Ile Glu Ala Gly Leu Trp Val
                335                 340                 345
Ser Leu Phe His Thr Ile Pro Pro Gly Met Pro Glu Phe Gln Ser
                350                 355                 360
Glu Gly His Cys Thr Leu Glu Cys Leu Glu Glu Ala Leu Glu Ala
                365                 370                 375
Glu Lys Pro Ser Gly Ile His Val Phe Ala Val Leu Leu His Ala
                380                 385                 390
His Leu Ala Gly Arg Gly Ile Arg Leu Arg His Phe Arg Lys Gly
                395                 400                 405
Lys Glu Met Lys Leu Leu Ala Tyr Asp Asp Asp Phe Asp Phe Asn
                410                 415                 420
Phe Gln Glu Phe Gln Tyr Leu Lys Glu Glu Gln Thr Ile Leu Pro
                425                 430                 435
Gly Asp Asn Leu Ile Thr Glu Cys Arg Tyr Asn Thr Lys Asp Arg
                440                 445                 450
Ala Glu Met Thr Trp Gly Gly Leu Ser Thr Arg Ser Glu Met Cys
                455                 460                 465
Leu Ser Tyr Leu Leu Tyr Tyr Pro Arg Ile Asn Leu Thr Arg Cys
                470                 475                 480
Ala Ser Ile Pro Asp Ile Met Glu Gln Leu Gln Phe Ile Gly Val
                485                 490                 495
```

```
Lys Glu Ile Tyr Arg Pro Val Thr Thr Trp Pro Phe Ile Ile Lys
                500                 505                 510
Ser Pro Lys Gln Tyr Lys Asn Leu Ser Phe Met Asp Ala Met Asn
                515                 520                 525
Lys Phe Lys Trp Thr Lys Lys Glu Gly Leu Ser Phe Asn Lys Leu
                530                 535                 540
Val Leu Ser Leu Pro Val Asn Val Arg Cys Ser Lys Thr Asp Asn
                545                 550                 555
Ala Glu Trp Ser Ile Gln Gly Met Thr Ala Leu Pro Pro Asp Ile
                560                 565                 570
Glu Arg Pro Tyr Lys Ala Glu Pro Leu Val Cys Gly Thr Ser Ser
                575                 580                 585
Ser Ser Ser Leu His Arg Asp Phe Ser Ile Asn Leu Leu Val Cys
                590                 595                 600
Leu Leu Leu Leu Ser Cys Thr Leu Ser Thr Lys Ser Leu
                605                 610
```

<210> 19
<211> 741
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4169101CD1

<400> 19

```
Met Ala Val Leu Asp Thr Asp Leu Asp His Ile Leu Pro Ser Ser
  1                 5                  10                  15
Val Leu Pro Pro Phe Trp Ala Lys Leu Val Val Gly Ser Val Ala
                 20                  25                  30
Ile Val Cys Phe Ala Arg Ser Tyr Asp Gly Asp Phe Val Phe Asp
                 35                  40                  45
Asp Ser Glu Ala Ile Val Asn Asn Lys Asp Leu Gln Ala Glu Thr
                 50                  55                  60
Pro Leu Gly Asp Leu Trp His His Asp Phe Trp Gly Ser Arg Leu
                 65                  70                  75
Ser Ser Asn Thr Ser His Lys Ser Tyr Arg Pro Leu Thr Val Leu
                 80                  85                  90
Thr Phe Arg Ile Asn Tyr Tyr Leu Ser Gly Gly Phe His Pro Val
                 95                 100                 105
Gly Phe His Val Val Asn Ile Leu Leu His Ser Gly Ile Ser Val
                110                 115                 120
Leu Met Val Asp Val Phe Ser Val Leu Phe Gly Gly Leu Gln Tyr
                125                 130                 135
Thr Ser Lys Gly Arg Arg Leu His Leu Ala Pro Arg Ala Ser Leu
                140                 145                 150
Leu Ala Ala Leu Leu Phe Ala Val His Pro Val His Thr Glu Cys
                155                 160                 165
Val Ala Gly Val Val Gly Arg Ala Asp Leu Leu Cys Ala Leu Phe
                170                 175                 180
Phe Leu Leu Ser Phe Leu Gly Tyr Cys Lys Ala Phe Arg Glu Ser
                185                 190                 195
Asn Lys Glu Gly Ala His Ser Ser Thr Phe Trp Val Leu Leu Ser
                200                 205                 210
Ile Phe Leu Gly Ala Val Ala Met Leu Cys Lys Glu Gln Gly Ile
                215                 220                 225
Thr Val Leu Gly Leu Asn Ala Val Phe Asp Ile Leu Val Ile Gly
                230                 235                 240
Lys Phe Asn Val Leu Glu Ile Val Gln Lys Val Leu His Lys Asp
                245                 250                 255
Lys Ser Leu Glu Asn Leu Gly Met Leu Arg Asn Gly Gly Leu Leu
                260                 265                 270
Phe Arg Met Thr Leu Leu Thr Ser Gly Gly Ala Gly Met Leu Tyr
```

```
                275                 280                 285
Val Arg Trp Arg Ile Met Gly Thr Gly Pro Pro Ala Phe Thr Glu
            290                 295                 300
Val Asp Asn Pro Ala Ser Phe Ala Asp Ser Met Leu Val Arg Ala
        305                 310                 315
Val Asn Tyr Asn Tyr Tyr Tyr Ser Leu Asn Ala Trp Leu Leu Leu
    320                 325                 330
Cys Pro Trp Trp Leu Cys Phe Asp Trp Ser Met Gly Cys Ile Pro
                335                 340                 345
Leu Ile Lys Ser Ile Ser Asp Trp Arg Val Ile Ala Leu Ala Ala
            350                 355                 360
Leu Trp Phe Cys Leu Ile Gly Leu Ile Cys Gln Ala Leu Cys Ser
        365                 370                 375
Glu Asp Gly His Lys Arg Arg Ile Leu Thr Leu Gly Leu Gly Phe
    380                 385                 390
Leu Val Ile Pro Phe Leu Pro Ala Ser Asn Leu Phe Phe Arg Val
                395                 400                 405
Gly Phe Val Val Ala Glu Arg Val Leu Tyr Leu Pro Ser Ile Gly
            410                 415                 420
Tyr Cys Val Leu Leu Thr Phe Gly Phe Gly Ala Leu Ser Lys His
        425                 430                 435
Thr Lys Lys Lys Lys Leu Ile Ala Ala Val Val Leu Gly Ile Leu
    440                 445                 450
Phe Ile Asn Thr Leu Arg Cys Val Leu Arg Ser Gly Glu Trp Arg
                455                 460                 465
Ser Glu Glu Gln Leu Phe Arg Ser Ala Leu Ser Val Cys Pro Leu
            470                 475                 480
Asn Ala Lys Val His Tyr Asn Ile Gly Lys Asn Leu Ala Asp Lys
        485                 490                 495
Gly Asn Gln Thr Ala Ala Ile Arg Tyr Tyr Arg Glu Ala Val Arg
    500                 505                 510
Leu Asn Pro Lys Tyr Val His Ala Met Asn Asn Leu Gly Asn Ile
                515                 520                 525
Leu Lys Glu Arg Asn Glu Leu Gln Glu Ala Glu Glu Leu Leu Ser
            530                 535                 540
Leu Ala Val Gln Ile Gln Pro Asp Phe Ala Ala Ala Trp Met Asn
        545                 550                 555
Leu Gly Ile Val Gln Asn Ser Leu Lys Arg Phe Glu Ala Ala Glu
    560                 565                 570
Gln Ser Tyr Arg Thr Ala Ile Lys His Arg Arg Lys Tyr Pro Asp
                575                 580                 585
Cys Tyr Tyr Asn Leu Gly Arg Leu Tyr Ala Asp Leu Asn Arg His
            590                 595                 600
Val Asp Ala Leu Asn Ala Trp Arg Asn Ala Thr Val Leu Lys Pro
        605                 610                 615
Glu His Ser Leu Ala Trp Asn Asn Met Ile Ile Leu Leu Asp Asn
    620                 625                 630
Thr Gly Asn Leu Ala Gln Ala Glu Ala Val Gly Arg Glu Ala Leu
                635                 640                 645
Glu Leu Ile Pro Asn Asp His Ser Leu Met Phe Ser Leu Ala Asn
            650                 655                 660
Val Leu Gly Lys Ser Gln Lys Tyr Lys Glu Ser Glu Ala Leu Phe
        665                 670                 675
Leu Lys Ala Ile Lys Ala Asn Pro Asn Ala Ala Ser Tyr His Gly
    680                 685                 690
Asn Leu Ala Val Leu Tyr His Arg Trp Gly His Leu Asp Leu Ala
                695                 700                 705
Lys Lys His Tyr Glu Ile Ser Leu Gln Leu Asp Pro Thr Ala Ser
            710                 715                 720
Gly Thr Lys Glu Asn Tyr Gly Leu Leu Arg Arg Lys Leu Glu Leu
        725                 730                 735
Met Gln Lys Lys Ala Val
    740
```

<210> 20
<211> 535
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2925182CD1

<400> 20

```
Met Arg Leu Arg Asn Gly Thr Val Ala Thr Ala Leu Ala Phe Ile
  1               5                  10                  15
Thr Ser Phe Leu Thr Leu Ser Trp Tyr Thr Thr Trp Gln Asn Gly
             20                  25                  30
Lys Glu Lys Leu Ile Ala Tyr Gln Arg Glu Phe Leu Ala Leu Lys
             35                  40                  45
Glu Arg Leu Arg Ile Ala Glu His Arg Ile Ser Gln Arg Ser Ser
             50                  55                  60
Glu Leu Asn Thr Ile Val Gln Gln Phe Lys Arg Val Gly Ala Glu
             65                  70                  75
Thr Asn Gly Ser Lys Asp Ala Leu Asn Lys Phe Ser Asp Asn Thr
             80                  85                  90
Leu Lys Leu Leu Lys Glu Leu Thr Ser Lys Lys Ser Leu Gln Val
             95                 100                 105
Pro Ser Ile Tyr Tyr His Leu Pro His Leu Leu Lys Asn Glu Gly
            110                 115                 120
Ser Leu Gln Pro Ala Val Gln Ile Gly Asn Gly Arg Thr Gly Val
            125                 130                 135
Ser Ile Val Met Gly Ile Pro Thr Val Lys Arg Glu Val Lys Ser
            140                 145                 150
Tyr Leu Ile Glu Thr Leu His Ser Leu Ile Asp Asn Leu Tyr Pro
            155                 160                 165
Glu Glu Lys Leu Asp Cys Val Ile Val Val Phe Ile Gly Glu Thr
            170                 175                 180
Asp Ile Asp Tyr Val His Gly Val Val Ala Asn Leu Glu Lys Glu
            185                 190                 195
Phe Ser Lys Glu Ile Ser Ser Gly Leu Val Glu Val Ile Ser Pro
            200                 205                 210
Pro Glu Ser Tyr Tyr Pro Asp Leu Thr Asn Leu Lys Glu Thr Phe
            215                 220                 225
Gly Asp Ser Lys Glu Arg Val Arg Trp Arg Thr Lys Gln Asn Leu
            230                 235                 240
Asp Tyr Cys Phe Leu Met Met Tyr Ala Gln Glu Lys Gly Ile Tyr
            245                 250                 255
Tyr Ile Gln Leu Glu Asp Asp Ile Ile Val Lys Gln Asn Tyr Phe
            260                 265                 270
Asn Thr Ile Lys Asn Phe Ala Leu Gln Leu Ser Ser Glu Glu Trp
            275                 280                 285
Met Ile Leu Glu Phe Ser Gln Leu Gly Phe Ile Gly Lys Met Phe
            290                 295                 300
Gln Ala Pro Asp Leu Thr Leu Ile Val Glu Phe Ile Phe Met Phe
            305                 310                 315
Tyr Lys Glu Lys Pro Ile Asp Trp Leu Leu Asp His Ile Leu Trp
            320                 325                 330
Val Lys Val Cys Asn Pro Glu Lys Asp Ala Lys His Cys Asp Arg
            335                 340                 345
Gln Lys Ala Asn Leu Arg Ile Arg Phe Arg Pro Ser Leu Phe Gln
            350                 355                 360
His Val Gly Leu His Ser Ser Leu Ser Gly Lys Ile Gln Lys Leu
            365                 370                 375
Thr Asp Lys Asp Tyr Met Lys Pro Leu Leu Leu Lys Ile His Val
            380                 385                 390
Asn Pro Pro Ala Glu Val Ser Thr Ser Leu Lys Val Tyr Gln Gly
            395                 400                 405
```

```
His Thr Leu Glu Lys Thr Tyr Met Gly Glu Asp Phe Phe Trp Ala
                410                 415                 420
Ile Thr Pro Ile Ala Gly Asp Tyr Ile Leu Phe Lys Phe Asp Lys
                425                 430                 435
Pro Val Asn Val Glu Ser Tyr Leu Phe His Ser Gly Asn Gln Glu
                440                 445                 450
His Pro Gly Asp Ile Leu Leu Asn Thr Thr Val Glu Val Leu Pro
                455                 460                 465
Phe Lys Ser Glu Gly Leu Glu Ile Ser Lys Glu Thr Lys Asp Lys
                470                 475                 480
Arg Leu Glu Asp Gly Tyr Phe Arg Ile Gly Lys Phe Glu Asn Gly
                485                 490                 495
Val Ala Glu Gly Met Val Asp Pro Ser Leu Asn Pro Ile Ser Ala
                500                 505                 510
Phe Arg Leu Ser Val Ile Gln Asn Ser Ala Val Trp Ala Ile Leu
                515                 520                 525
Asn Glu Ile His Ile Lys Lys Ala Thr Asn
                530                 535

<210> 21
<211> 522
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3271838CD1

<400> 21
Met Ala Ala Met Ala Val Ala Leu Arg Gly Leu Gly Gly Arg Phe
  1                 5                  10                  15
Arg Trp Arg Thr Gln Ala Val Ala Gly Gly Val Arg Gly Ala Ala
                 20                  25                  30
Arg Gly Ala Ala Ala Gly Gln Arg Asp Tyr Asp Leu Leu Val Val
                 35                  40                  45
Gly Gly Gly Ser Gly Gly Leu Ala Cys Ala Lys Glu Ala Ala Gln
                 50                  55                  60
Leu Gly Arg Lys Val Ser Val Val Asp Tyr Val Glu Pro Ser Pro
                 65                  70                  75
Gln Gly Thr Arg Trp Gly Leu Gly Gly Thr Cys Val Asn Val Gly
                 80                  85                  90
Cys Ile Pro Lys Lys Leu Met His Gln Ala Ala Leu Leu Gly Gly
                 95                 100                 105
Leu Ile Gln Asp Ala Pro Asn Tyr Gly Trp Glu Val Ala Gln Pro
                110                 115                 120
Val Pro His Asp Trp Arg Lys Met Ala Glu Ala Val Gln Asn His
                125                 130                 135
Val Lys Ser Leu Asn Trp Gly His Arg Val Gln Leu Gln Asp Arg
                140                 145                 150
Lys Val Lys Tyr Phe Asn Ile Lys Ala Ser Phe Val Asp Glu His
                155                 160                 165
Thr Val Cys Gly Val Ala Lys Gly Gly Lys Glu Ile Leu Leu Ser
                170                 175                 180
Ala Asp His Ile Ile Ile Ala Thr Gly Gly Arg Pro Arg Tyr Pro
                185                 190                 195
Thr His Ile Glu Gly Ala Leu Glu Tyr Gly Ile Thr Ser Asp Asp
                200                 205                 210
Ile Phe Trp Leu Lys Glu Ser Pro Gly Lys Thr Leu Val Val Gly
                215                 220                 225
Ala Ser Tyr Val Ala Leu Glu Cys Ala Gly Phe Leu Thr Gly Ile
                230                 235                 240
Gly Leu Asp Thr Thr Ile Met Met Arg Ser Ile Pro Leu Arg Gly
                245                 250                 255
Phe Asp Gln Gln Met Ser Ser Met Val Ile Glu His Met Ala Ser
```

```
                260                 265                 270
His Gly Thr Arg Phe Leu Arg Gly Cys Ala Pro Ser Arg Val Arg
                275                 280                 285
Arg Leu Pro Asp Gly Gln Leu Gln Val Thr Trp Glu Asp Arg Thr
                290                 295                 300
Thr Gly Lys Glu Asp Thr Gly Thr Phe Asp Thr Val Leu Trp Ala
                305                 310                 315
Ile Gly Arg Val Pro Asp Thr Arg Ser Leu Asn Leu Glu Lys Ala
                320                 325                 330
Gly Val Asp Thr Ser Pro Asp Thr Gln Lys Ile Leu Val Asp Ser
                335                 340                 345
Arg Glu Ala Thr Ser Val Pro His Ile Tyr Ala Ile Gly Asp Val
                350                 355                 360
Val Glu Gly Arg Pro Glu Leu Thr Pro Thr Ala Ile Met Ala Gly
                365                 370                 375
Arg Leu Leu Val Gln Arg Leu Phe Gly Gly Ser Ser Asp Leu Met
                380                 385                 390
Asp Tyr Asp Asn Val Pro Thr Thr Val Phe Thr Pro Leu Glu Tyr
                395                 400                 405
Gly Cys Val Gly Leu Ser Glu Glu Glu Ala Val Ala Arg His Gly
                410                 415                 420
Gln Glu His Val Glu Val Tyr His Ala His Tyr Lys Pro Leu Glu
                425                 430                 435
Phe Thr Val Ala Gly Arg Asp Ala Ser Gln Cys Tyr Val Lys Met
                440                 445                 450
Val Cys Leu Arg Glu Pro Pro Gln Leu Val Leu Gly Leu His Phe
                455                 460                 465
Leu Gly Pro Asn Ala Gly Glu Val Thr Gln Gly Phe Ala Leu Gly
                470                 475                 480
Ile Lys Cys Gly Ala Ser Tyr Ala Gln Val Met Arg Thr Val Gly
                485                 490                 495
Ile His Pro Thr Cys Ser Glu Glu Val Val Lys Leu Arg Ile Ser
                500                 505                 510
Lys Arg Ser Gly Leu Asp Pro Thr Val Thr Gly Cys
                515                 520
```

<210> 22
<211> 495
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3292871CD1

<400> 22

```
Met Lys Asn Lys Thr Cys Val Leu Val Cys Val Ser Val Phe Gly
  1                 5                  10                  15
Gly Glu Arg Gly Gln Val Thr Val Pro Arg Val Gly Val Arg Arg
                 20                  25                  30
Pro Ser Leu Ala Gly Pro Leu Gln Lys Cys Thr Leu Arg Glu Thr
                 35                  40                  45
Arg Val Trp Leu Pro Gln Gly Ser Gly Phe Gln Ser Ser Arg Arg
                 50                  55                  60
Glu Lys Tyr Gly Asn Val Phe Lys Thr His Leu Leu Gly Arg Pro
                 65                  70                  75
Leu Ile Arg Val Thr Gly Ala Glu Asn Val Arg Lys Ile Leu Met
                 80                  85                  90
Gly Glu His His Leu Val Ser Thr Glu Trp Pro Arg Ser Thr Arg
                 95                 100                 105
Met Leu Leu Gly Pro Asn Thr Val Ser Asn Ser Ile Gly Asp Ile
                110                 115                 120
His Arg Asn Lys Arg Lys Val Phe Ser Lys Ile Phe Ser His Glu
                125                 130                 135
```

```
Ala Leu Glu Ser Tyr Leu Pro Lys Ile Gln Leu Val Ile Gln Asp
                140                 145                 150
Thr Leu Arg Ala Trp Ser Ser His Pro Glu Ala Ile Asn Val Tyr
                155                 160                 165
Gln Glu Ala Gln Lys Leu Thr Phe Arg Met Ala Ile Arg Val Leu
                170                 175                 180
Leu Gly Phe Ser Ile Pro Glu Glu Asp Leu Gly His Leu Phe Glu
                185                 190                 195
Val Tyr Gln Gln Phe Val Asp Asn Val Phe Ser Leu Pro Val Asp
                200                 205                 210
Leu Pro Phe Ser Gly Tyr Arg Arg Gly Ile Gln Ala Arg Gln Ile
                215                 220                 225
Leu Gln Lys Gly Leu Glu Lys Ala Ile Arg Glu Lys Leu Gln Cys
                230                 235                 240
Thr Gln Gly Lys Asp Tyr Leu Asp Ala Leu Asp Leu Leu Ile Glu
                245                 250                 255
Ser Ser Lys Glu His Gly Lys Glu Met Thr Met Gln Glu Leu Lys
                260                 265                 270
Asp Gly Thr Leu Glu Leu Ile Phe Ala Ala Tyr Ala Thr Thr Ala
                275                 280                 285
Ser Ala Ser Thr Ser Leu Ile Met Gln Leu Leu Lys His Pro Thr
                290                 295                 300
Val Leu Glu Lys Leu Arg Asp Glu Leu Arg Ala His Gly Ile Leu
                305                 310                 315
His Ser Gly Gly Cys Pro Cys Glu Gly Thr Leu Arg Leu Asp Thr
                320                 325                 330
Leu Ser Gly Leu Arg Tyr Leu Asp Cys Val Ile Lys Glu Val Met
                335                 340                 345
Arg Leu Phe Thr Pro Ile Ser Gly Gly Tyr Arg Thr Val Leu Gln
                350                 355                 360
Thr Phe Glu Leu Asp Gly Phe Gln Ile Pro Lys Gly Trp Ser Val
                365                 370                 375
Met Tyr Ser Ile Arg Asp Thr His Asp Thr Ala Pro Val Phe Lys
                380                 385                 390
Asp Val Asn Val Phe Asp Pro Asp Arg Phe Ser Gln Ala Arg Ser
                395                 400                 405
Glu Asp Lys Asp Gly Arg Phe His Tyr Leu Pro Phe Gly Gly Gly
                410                 415                 420
Val Arg Thr Cys Leu Gly Lys His Leu Ala Lys Leu Phe Leu Lys
                425                 430                 435
Val Leu Ala Val Glu Leu Ala Ser Thr Ser Arg Phe Glu Leu Ala
                440                 445                 450
Thr Arg Thr Phe Pro Arg Ile Thr Leu Val Pro Val Leu His Pro
                455                 460                 465
Val Asp Gly Leu Ser Val Lys Phe Phe Gly Leu Asp Ser Asn Gln
                470                 475                 480
Asn Glu Ile Leu Pro Glu Thr Glu Ala Met Leu Ser Ala Thr Val
                485                 490                 495
```

<210> 23
<211> 51
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4109179CD1

<400> 23

```
Met Glu Glu Lys Thr Ile Leu Ser Cys Ile Leu Arg His Phe Trp
  1                 5                  10                  15
Ile Glu Ser Asn Gln Lys Arg Glu Glu Leu Gly Leu Glu Gly Gln
                 20                  25                  30
```

Leu Ile Leu Arg Pro Ser Asn Gly Ile Trp Ile Lys Leu Lys Arg
35 40 45
Arg Asn Ala Asp Glu Arg
50

<210> 24
<211> 335
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4780365CD1

<400> 24
Met Ile Leu Phe Leu Ile Met Leu Val Leu Val Leu Phe Gly Tyr
1 5 10 15
Gly Val Leu Ser Pro Arg Ser Leu Met Pro Gly Ser Leu Glu Arg
20 25 30
Gly Phe Cys Met Ala Val Arg Glu Pro Asp His Leu Gln Arg Val
35 40 45
Ser Leu Pro Arg Met Val Tyr Pro Gln Pro Lys Val Leu Thr Pro
50 55 60
Cys Arg Lys Asp Val Leu Val Val Thr Pro Trp Leu Ala Pro Ile
65 70 75
Val Trp Glu Gly Thr Phe Asn Ile Asp Ile Leu Asn Glu Gln Phe
80 85 90
Arg Leu Gln Asn Thr Thr Ile Gly Leu Thr Val Phe Ala Ile Lys
95 100 105
Lys Tyr Val Ala Phe Leu Lys Leu Phe Leu Glu Thr Ala Glu Lys
110 115 120
His Phe Met Val Gly His Arg Val His Tyr Tyr Val Phe Thr Asp
125 130 135
Gln Pro Ala Ala Val Pro Arg Val Thr Leu Gly Thr Gly Arg Gln
140 145 150
Leu Ser Val Leu Glu Val Arg Ala Tyr Lys Arg Trp Gln Asp Val
155 160 165
Ser Met Arg Arg Met Glu Met Ile Ser Asp Phe Cys Glu Arg Arg
170 175 180
Phe Leu Ser Glu Val Asp Tyr Leu Val Cys Val Asp Val Asp Met
185 190 195
Glu Phe Arg Asp His Val Gly Val Glu Ile Leu Thr Pro Leu Phe
200 205 210
Gly Thr Leu His Pro Gly Phe Tyr Gly Ser Ser Arg Glu Ala Phe
215 220 225
Thr Tyr Glu Arg Arg Pro Gln Ser Gln Ala Tyr Ile Pro Lys Asp
230 235 240
Glu Gly Asp Phe Tyr Tyr Leu Gly Gly Phe Phe Gly Gly Ser Val
245 250 255
Gln Glu Val Gln Arg Leu Thr Arg Ala Cys His Gln Ala Met Met
260 265 270
Val Asp Gln Ala Asn Gly Ile Glu Ala Val Trp His Asp Glu Ser
275 280 285
His Leu Asn Lys Tyr Leu Leu Arg His Lys Pro Thr Lys Val Leu
290 295 300
Ser Pro Glu Tyr Leu Trp Asp Gln Gln Leu Leu Gly Trp Pro Ala
305 310 315
Val Leu Arg Lys Leu Arg Phe Thr Ala Val Pro Lys Asn His Gln
320 325 330
Ala Val Arg Asn Pro
335

<210> 25
<211> 1269

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1799250CB1

<400> 25
cgcggcggcg gcgcggccgg ggcagccatg tcgccattgt ctgcagcgcg ggcggccctg 60
cgggtctacg cggtaggcgc cgcggtgatc ctggcgcagc tgctgcggcg ctgccgcggg 120
ggcttcctgg agccagttct ccccccacga cctgaccgtg tcgctatagt gacgggaggg 180
acagatggca ttggctattc tacagcgaag catctggcga gacttggcat gcatgttatc 240
atagctggaa ataatgacag caaagccaaa caagttgtaa gcaaaataaa agaagaaacc 300
ttgaacgaca aagtggaatt tttatactgt gacttggctt ccatgacttc catccggcag 360
tttgtgcaga agttcaagat gaagaagatt cctctccatg tcctgatcaa caatgctggg 420
gtgatgatgg tccctcagag gaaaaccaga gatggattcg aagaacattt cggcctgaac 480
tacctagggc acttcctgct gaccaacctt ctcttggata cgctgaaaga gtctgggtcc 540
cctggccaca gtgcgagggt ggtcaccgtc tcctctgcca cccattacgt cgctgagctg 600
aacatggatg accttcagag cagtgcctgc tactcacccc acgcagccta cgcccagagc 660
aagctggccc ttgtcctgtt cacctaccac ctccagcggc tgctggcggc tgagggaagc 720
cacgtgaccg ccaacgtggt ggaccccggg gtggtcaaca cggacgtcta caagcacgtg 780
ttctgggcca cccgtctggc gaagaagctt ctcggctggt tgcttttcaa gacccccgat 840
gaaggagcgt ggacttccat ctacgcagca gtcaccccag agctggaagg agttggtggc 900
cgttacctat acaacaagaa agagaccaag tccctccacg tcacctacaa ccagaaactg 960
cagcagcagc tgtggtctaa gagttgtgag atgactgggg tccttgatgt gaccctgtga 1020
tatcctgtct caggatagct gctgccccaa gaaacacatt gcacctgcca atagcttgtg 1080
ggtctgtgaa gactgcggtg tttgagtttc tcacacccac ctgcccacag ggctctgtcc 1140
tctagttttg agacagctgc ctcaacctct gcagaacttc aagaagccaa ataaacattt 1200
tggaggataa tcaccccaag tggtcttcaa ccataaactt tgtgattcca aagtgcccag 1260
ttgtcacag 1269

<210> 26
<211> 1593
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2242475CB1

<400> 26
cctgcctggt cctctgtgcc tggtggggtg ggggtgccag gtgtgtccag aggagcccat 60
ttggtagtga ggcaggtatg gggctagaag cactggtgcc cctggccgtg atagtggcca 120
tcttcctgct cctggtggac ctgatgcacc ggcgccaacg ctgggctgca cgctacccac 180
caggcccccc gccactgccc gggctgggca acctgctgca tgtggacttc cagaacacac 240
catactgctt cgaccagttg cggcgccgct tcggggacgt gttcagcctg cagctggcct 300
ggacgccggt ggtcgtgctc aatgggctgg cggccgtgcg cgaggcgctg gtgacccacg 360
gcgaggacac cgccgaccgc ccgcctgtgc ccatcaccca gatcctgggt ttcgggccgc 420
gttcccaagg ggtgttcctg gcgcgctatg ggcccgcgtg gcgcgagcag aggcgcttct 480
ccgtctccac cttgcgcaac ttgggcctgg gcaagaagtc gctggagcag tgggtgaccg 540
aggaggccgc ctgcctttgt gccgccttcg ccaaccactc cggacgcccc tttcgcccca 600
acggtctctt ggacaaagcc gtgagcaacg tgatcgcctc cctcacctgc gggcgccgct 660
tcgagtacga cgaccctcgc ttcctcaggc tgctggacct agctcaggag ggactgaagg 720
aggagtcggg cttcctgcgc gaggtgctga atgctgtccc cgtcctcccg cacatcccag 780
cgctggctgg caaggtccta cgcttccaaa aggctttcct gacccagctg gatgagctgc 840
taactgagca caggatgacc tgggacccag cccagccacc ccgagacctg actgaggcct 900
tcctggcaaa gaaggagaag gccaagggga gccctgagag cagcttcaat gatgagaacc 960
tgcgcatagt ggtgggtaac ctgttccttg ccgggatggt gaccacctcg accacactgg 1020
cctgggccct gctgctcatg atcctgcatc cggatgtgca gtgccgagta caacaggaaa 1080
tcgatgaggt catagggcag gtgcggcatc cagagatggc agaccaggcc cacatgccgt 1140
tcaccaatgc tgtcatccat gaggtgcagc gctttgcaga cattgtccca atgaatttgc 1200
cacacaagac ttctcgtgac attgaagtgc agggcttcct tatccctaag gggacaaccc 1260
tcatccccaa cctgtcctca gtgctgaagg atgagactgt ctgggagaag cccctccgat 1320
tccaccctga acacttcctg gatgcccagg gcaactttgt gaagcatgag gccttcatgc 1380

cattctcagc aggccgcaga gcatgcctgg gggagcccct ggcccgcatg gagctcttcc 1440
tcttcttcac ctgcctcctg caacgcttca gcttctccgt gcccactgga cagccccggc 1500
ccagcgacta tggtgtcttt gcctttctcc ttagcccttc cccctaccag ctctgtgcat 1560
tcaaacgtta gaaggaaaga aattctagtc cag 1593

<210> 27
<211> 1779
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2706492CB1

<400> 27
cagtgagaga actgagaccc agagagatta agtatcttgc ccaaggtcac actttagtaa 60
aaggcaaagt caggatttga atccacacac ttatctagta cactctaaga cacaggggca 120
gattttagta aacagtagga gatggactct cagaatttgg tgcctgggga gagggaagag 180
gagagagatg cctggtgaca agcccagccc ttgcctctcc acaggagagt gggaacaatg 240
aggtcatctt catggccttg gacttggcca gtctggcctc ggtgcgggcc tttgccactg 300
cctttctgag ctctgagcca cggttggaca tcctcatcca caatgccggt atcagttcct 360
gtggccggac ccgtgaggcg tttaacctgc tgcttcgggt gaaccatatc ggtccctttc 420
tgctgacaca tctgctgctg ccttgcctga aggcatgtgc ccctagccgc gtggtggtgg 480
tagcctcagc tgcccactgt cggggacgtc ttgacttcaa acgcctggac cgcccagtgg 540
tgggctggcg gcaggagctg cgggcatatg ctgacactaa gctggctaat gtactgtttg 600
cccgggagct cgccaaccag cttgaggcca ctggcgtcac ctgctatgca gcccacccag 660
ggcctgtgaa ctcggagctg ttcctgcgcc atgttcctgg atggctgcgc ccacttttgc 720
gcccattggc ttggctggtg ctccgggcac caagaggggg tgcccagaca cccctgtatt 780
gtgctctaca agagggcatc gagcccctca gtgggagata ttttgccaac tgccatgtgg 840
aagaggtgcc tccagctgcc cgagacgacc gggcagccca tcggctatgg gaggccagca 900
agaggctggc agggcttggg cctggggagg atgctgaacc cgatgaagac ccccagtctg 960
aggactcaga ggccccatct tctctaagca ccccccaccc tgaggagccc acagtttctc 1020
aaccttaccc cagccctcag agctcaccag atttgtctaa gatgacgcac cgaattcagg 1080
ctaaagttga gcctgagatc cagctctcct aaccctcagg ccaggatgct tgccatggca 1140
cttcatggtc cttgaaaacc tcggatgtgt gcgaggccat gccctggaca ctgacgggtt 1200
tgtgatcttg acctccgtgg ttactttctg gggccccaag ctgtgccctg gacatctctt 1260
ttcctggttg aaggaataat gggtgattat ttcttcctga gagtgacagt aaccccagat 1320
ggagagatag gggtatgcta gacactgtgc ttctcggaaa tttggatgta gtattttcag 1380
gccccaccct tattgattct gatcagctct ggagcagagg cagggagttt gcaatgtgat 1440
gcactgccaa cattgagaat tagtgaactg atccctttgc aaccgtctag ctaggtagtt 1500
aaattacccc catgttaatg aagcggaatt aggctcccga gctaagggac tcgcctaggg 1560
tctcacagtg agtaggagga gggcctggga tctgaaccca agggtctgag gccagggccg 1620
actgccgtaa gatgggtgct gagaagtgag tcagggcagg gcagctggta tcgaggtgcc 1680
ccatgggagt aaggggacgc cttccgggcg gatgcagggc tggggtcatc tgtatctgaa 1740
gcccctcgga ataaagcgcg ttgaccgccg aaaaaaaaa 1779

<210> 28
<211> 1931
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2766688CB1

<400> 28
ggcaacgcgg ctggttctcg cccgtcagtc ctagcccggc cctgcccctc gcttgcattt 60
tttccgcgct ggctgagatt caaagagaag tggaggtggg agggagcgac aatggaaaaa 120
tcacctgaaa actgggacag aggaaggaag ctacagttac gaaggagagc tgcaaaagtt 180
gcagcagaaa ggttgggagt cccgacaggt tccgtagccc acagaaaaga agcaagggac 240
ggcaggactg tttcacactt ttctgcttct ggaaggtgct ggacaaaaac atggaactaa 300
tttccccaac agtgattata atcctgggtt gccttgctct gttcttactc cttcagcgga 360
agaatttgcg tagacccccg tgcatcaagg gctggattcc ttggattgga gttggatttg 420
agtttgggaa agcccctcta gaatttatag agaaagcaag aatcaagtat ggaccaatat 480

ttacagtctt tgctatggga aaccgaatga cctttgttac tgaagaagaa ggaattaatg 540
tgtttctaaa atccaaaaaa gtagattttg aactagcagt gcaaaatatc gtttatcata 600
cagcatcaat tccaaagaat gtctttttag cactgcatga aaaactctat attatgttga 660
aagggaaaat ggggactgtc aatctccatc agtttactgg gcaactgact gaagaattac 720
atgaacaact ggagaattta ggcactcatg ggacaatgga cctgaacaac ttagtaagac 780
atctccttta tccagtcaca gtgaatatgc tctttaataa aagtttgttt tccacaaaca 840
agaaaaaaat caaggagttc catcagtatt ttcaagttta tgatgaagat tttgagtatg 900
ggtcccagtt gccagagtgt cttctaagaa actggtcaaa atccaaaaag tggttcctgg 960
aactgtttga gaaaaacatt ccagatataa aagcatgtaa atctgcaaaa gataattcca 1020
tgacattatt gcaagctacg ctggatattg tagagacgga aacaagtaag gaaaactcac 1080
ccaattatgg gctcttactg ctttgggctt ctctgtctaa tgctgttcct gttgcatttt 1140
ggacacttgc atacgtcctt tctcatcctg atatccacaa ggccattatg gaaggcatat 1200
cttctgtgtt tggcaaagca ggcaaagata agattaaagt gtctgaggat gacctggaga 1260
atctccttct aattaaatgg tgtgttttgg aaaccattcg tttaaaagct cctggtgtca 1320
ttactagaaa agtggtgaag cctgtggaaa ttttgaatta catcattcct tctggtgact 1380
tgttgatgtt gtctccattt tggctgcata gaaatccaaa gtattttcct gagcctgaat 1440
tgttcaaacc tgaacgttgg gaaaaaggca aatttagaga agcactcttt cttggactgc 1500
ttcatggcat tggaagcggg aagttccagt gtcctgcaag gtggtttgct ctgttagagg 1560
ttcagatgtg tattatttta atactttata aatatgactg tagtcttctg gacccattac 1620
ccaaacagag ttatctccat ttggtgggtg tcccccagcc ggaagggcaa tgccgaattg 1680
aatataaaca aagaatatga catctgttgg gcctcacaag gaccagggcc ttctggagga 1740
gtggcactac cccacctggc agcacctaga cctgagctct acaaaaacac actgcttcac 1800
tttgttttag gacttagttc aagaacacat tcaaatggtg catgtgtttg gtatcttcaa 1860
cagtagacca agaatctaac atcactctca gtaatataga gaccggaata catggtttat 1920
aggaaatgat c 1931

<210> 29
<211> 1282
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2788823CB1

<400> 29
cgcgcctgcg cctccgctcg cctgtggctg cgtcgcgcgc tcttcctcgg agctacccag 60
gcggctggtg tgcagcaagc tccgcgccga ccccggacgc ctgacgcctg acgcctgtcc 120
ccggcccggc atgagccgct acctgctgcc gctgtcggcg ctgggcacgg tagcaggcgc 180
cgccgtgctg ctcaaggact atgtcaccgg tggggcttgc cccagcaagg ccaccatccc 240
tgggaagacg gtcatcgtga cgggcgccaa cacaggcatc gggaagcaga ccgccttgga 300
actggccagg agaggaggca acatcatcct ggcctgccga gacatggaga agtgtgaggc 360
ggcagcaaag gacatccgcg gggagaccct caatcaccat gtcaacgccc ggcacctgga 420
cttggcttcc ctcaagtcta tccgagagtt tgcagcaaag atcattgaag aggaggagcg 480
agtggacatt ctaatcaaca acgcgggtgt gatgcggtgc ccccactgga ccaccgagga 540
cggcttcgag atgcagtttg gcgttaacca cctgggtcac tttctcttga caaacttgct 600
gctggacaag ctgaaagcct cagccccttc gcggatcatc aacctctcgt ccctggccca 660
tgttgctggg cacatagact ttgacgactt gaactggcag acgaggaagt ataacaccaa 720
agccgcctac tgccagagca agctcgccat cgtcctcttc accaaggagc tgagccggcg 780
gctgcaaggc tctggtgtga ctgtcaacgc cctgcacccc ggcgtggcca ggacagagct 840
gggcagacac acgggcatcc atggctccac cttctccagc accacactcg ggcccatctt 900
ctggctgctg gtcaagagcc ccgagctggc cgcccagccc agcacatacc tggccgtggc 960
ggaggaactg gcggatgttt ccggaaagta cttcgatgga ctcaaacaga aggccccggc 1020
ccccgaggct gaggatgagg aggtggcccg gaggctttgg gctgaaagtg cccgcctggt 1080
gggcttagag gctccctctg tgagggagca gcccctcccc agataacctc tggagcagat 1140
ttgaaagcca ggatggcgcc tccagaccga ggacagctgt ccgccatgcc cgcagcttcc 1200
tggcactacc tgagccggga gacccaggac tggcggccgc catgcccgca gtaggttcta 1260
gggggcggtg ctggccgcag tg 1282

<210> 30
<211> 2416
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3348822CB1

<400> 30
agcgtgcgcg ctttggtaac cggctagaaa tcccgcacgc gcgcctgcct cctctcccca 60
ggcctgagct gcccctccca ctgcctttcc ttcttcccgc gagtcagaag cttcgcgagg 120
gcccagagag gcggtggggt gggcgaccct acgccagctc cgggcgggag aaagcccacc 180
ctctcccgcg ccccaggaaa ccgccggcgt tcggcgctgc gcagagccat ggaattctcc 240
tggctggaga cgcgctgggc gcggcccttt tacctggcgt tcgtgttctg cctggccctg 300
gggctgctgc aggccattaa gctgtacctg cggaggcagc ggctgctgcg ggacctgcgc 360
cccttcccag cgcccccccac ccactggttc cttgggcacc agaagtttat tcaggatgat 420
aacatggaga agcttgagga aattattgaa aaataccctc gtgccttccc tttctggatt 480
gggccctttc aggcattttt ctgtatctat gacccagact atgcaaagac acttctgagc 540
agaacagatc ccaagtccca gtacctgcag aaattctcac ctccacttct tggaaaagga 600
ctagcggctc tagacggacc caagtggttc cagcatcgtc gcctactaac tcctggattc 660
cattttaaca tcctgaaagc atacattgag gtgatggctc attctgtgaa aatgatgctg 720
gataagtggg agaagatttg cagcactcag gacacaagcg tggaggtcta tgagcacatc 780
aactcgatgt ctctggatat aatcatgaaa tgcgctttca gcaaggagac caactgccag 840
acaaacagca cccatgatcc ttatgcaaaa gccatatttg aactcagcaa aatcatattt 900
caccgcttgt acagtttgtt gtatcacagt gacataattt tcaaactcag ccctcagggc 960
taccgcttcc agaagttaag ccgagtgttg aatcagtaca cagatacaat aatccaggaa 1020
agaaagaaat ccctccaggc tggggtaaag caggataaca ctccgaagag gaagtaccag 1080
gattttctgg atattgtcct ttctgccaag gatgaaagtg gtagcagctt ctcagatatt 1140
gatgtacact ctgaagtgag cacattcctg ttggcaggac atgacacctt ggcagcaagc 1200
atctcctgga tcctttactg cctggctctg aaccctgagc atcaagagag atgccgggag 1260
gaggtcaggg gcatcctggg ggatgggtct tctatcactt gggaccagct gggtgagatg 1320
tcgtacacca caatgtgcat caaggagacg tgccgattga ttcctgcagt cccgtccatt 1380
tccagagatc tcagcaagcc acttaccttc ccagatggat gcacattgcc tgcagggatc 1440
accgtggttc ttagtatttg gggtcttcac cacaaccctg ctgtctggaa aaacccaaag 1500
gtctttgacc ccttgaggtt ctctcaggag aattctgatc agagacaccc ctatgcctac 1560
ttaccattct cagctggatc aaggaactgc attgggcagg agtttgccat gattgagtta 1620
aaggtaacca ttgccttgat tctgctccac ttcagagtga ctccagaccc caccaggcct 1680
cttactttcc ccaaccattt tatcctcaag cccaagaatg ggatgtattt gcacctgaag 1740
aaactctctg aatgttagat ctcagggtac aatgattaaa cgtactttgt ttttcgaagt 1800
taaatttaca gctaatgatc caagcagata gaaagggatc aatgtatggt gggaggattg 1860
gaggttggtg ggataggggt ctctgtgaag agatccaaaa tcatttctag gtacacagtg 1920
tgtcagctag atctgtttct atataacttt gggagatttt cagatctttt ctgttaaact 1980
ttcactacta ttaatgctgt atacaccaat agactttcat atattttctg ttgtttttaa 2040
aatagttttc agaattatgc aagtaataag tgcatgtatg ctcactgtca aaaattccca 2100
acactagaaa atcatgtaga ataaaaattt taaatctcac ttcacttagc cgacattcca 2160
tgccctgacc aatcctactg cttttcctaa aaacagaata atttggtgtg cattctttca 2220
gacttttcc tatacatttt atatgtagaa atgtagcaat gtatttgtat agatgtgatc 2280
attcctatat tgttattgat ttttttcact taataaaaat tcaccttatt ccttatcatt 2340
gctttatggt attctgtaat atgaatgtac tataatttat ttaactattt tccttattgg 2400
gcatttaagt gatttc 2416

<210> 31
<211> 1574
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4290251CB1

<400> 31
ggacaacctg agtgctcagt cgtaaagagg aaaggcagaa tttttccttg ctatggctgg 60
aacaaacgca cttttgatgc tggaaaactt catagatgga aaatttttac cttgtagctc 120
atatatagat tcttacgacc catcaacagg ggaagtgtat tgcagagtgc caaatagtgg 180
aaaagacgag atcgaagccg cggtcaaggc cgccagagaa gcctttccca gctggtcatc 240
ccgcagcccc caggagcgct cacgggtcct gaaccaggtg gcggatttgc tggagcagtc 300
cctggaggag tttgcccagg ccgagtctaa agaccaaggg aaaaccttag cactggcaag 360
aaccatggac attccccggt ctgtgcagaa cttcaggttc ttcgcttcct ccagcctgca 420

ccacacgtca gagtgcacgc agatggaaca cctgggctgc atgcactaca cggtgcgggc 480
cccggtggga gtcgctggtc tgatcagccc ctggaatttg ccactctact tgctgacctg 540
gaagatagct ccagcgatgg ctgcagggaa cactgtgata gccaagccca gtgagctgac 600
ttcagtgact gcgtggatgt tgtgcaaact cctggataaa gcaggtgttc caccaggtgt 660
ggtcaatatt gtgtttggaa ccgggcccag ggtgggtgag gccctggtgt cccacccaga 720
ggtgcccctg atctccttca ccgggagcca gcccaccgct gagcggatca cccagctgag 780
cgctccccac tgcaaaaagc tctccctgga gctggggggc aagaatcctg ccatcatctt 840
tgaggacgcc aacctggatg agtgcattcc ggcaaccgtc aggtccagct ttgccaacca 900
ggtcagaagt tacgtcaaga gagctcttgc tgaaagtgcc caaatttggt gcggtgaagg 960
agtggataag ttgagcctcc ctgccaggaa ccaggcaggc tactttatgc ttcccacggt 1020
gataacagac attaaggatg aatcctgctg catgacggaa gagatatttg gtccagtgac 1080
gtgtgtcgtc ccctttgata gtgaagagga ggtgattgaa agagccaaca acgttaagta 1140
tgggctggcg gctaccgtgt ggtccagcaa tgtggggcgc gtccaccggg tggctaagaa 1200
gctgcagtct ggcttggtct ggaccaactg ctggctcatc agggagctga accttccttt 1260
cgggggggatg aagagttctg gaataggtag agagggagcc aaggactctt acgacttctt 1320
cactgagatc aaaaccatca ccgttaaaca ctgatctttg ctaatggtgg agccactatg 1380
gccaatgcct ggctgcaggc atcagttgtt caatgtggta gatgaaaatc atggcatgaa 1440
ttccagctat gccttgactt ggcagaaggt tatctctagc ttatcctcag ttcttagtaa 1500
ctttacccac tagtgaagag atactgtcta ttttcaatgt ggactcggaa aaaaagactt 1560
ataagtagga agat 1574

<210> 32
<211> 1227
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4904188CB1

<400> 32
gcaggaccca gggcgctgaa ctctcacaac caatcaggcg acccccccag agggaaacta 60
caagtcccag catgccccac gcgcaccgtc aggggccgac ccgccgcgcc ccagcgttct 120
ccgcgtacag gtggtctctt gggttccgga agagcctagg ctggatgtct tgatcaataa 180
cgcagggatc ttccagtgcc cttacatgaa gactgaagat gggtttgaga tgcagttcgg 240
agtgaaccat ctggggcact ttctactcac caatcttctc cttggactcc tcaaaagttc 300
agctcccagc aggattgtgg tagtttcttc caaactttat aaatacggag acatcaattt 360
tgatgacttg aacagtgaac aaagctataa taaaagcttt tgttatagcc ggagcaaact 420
ggctaacatt ctttttacca gggaactagc ccgccgctta gaaggcacaa atgtcaccgt 480
caatgtgttg catcctggta ttgtacggac aaatctgggg aggcacatac acattccact 540
gttggtcaaa ccactcttca atttggtgtc atgggctttt ttcaaaactc cagtagaagg 600
tgcccagact tccatttatt tggcctcttc acctgaggta gaaggagtgt caggaagata 660
ctttggggat tgtaaagagg aagaactgtt gcccaaagct atggatgaat ctgttgcaag 720
aaaactctgg gatatcagtg aagtgatggt tggcctgcta aaataggaac aaggagtaaa 780
agagctgttt ataaaactgc atatcagtta tatctgtgat caggaatggt gtggattgag 840
aacttgttac ttgaagaaaa agaattttga tattggaata gcctgctaag aggtacatgt 900
gggtattttg gagttactga aaaattattt ttgggataag agaatttcag caaagatgtt 960
ttaaatatat atagtaagta taatgaataa taagtacaat gaaaaataca attatattgt 1020
aaaattataa ctgggcaagc atggatgaca tattaatatt tgtcagaatt aagtgactca 1080
aagtgctatc gagaggtttt tcaagtatct ttgagtttca tggccaaagt gttaactagt 1140
tttactacaa tgtttggtgt ttgtgtggaa attatctgcc tggtgtgtgc acacaagtct 1200
tacttggaat aaatttactg gtacaaa 1227

<210> 33
<211> 1240
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 638419CB1

<400> 33
cttgcttgca cagtgtcctg gagctggacc tggctctggg tttccaggaa gcagtttgac 60

taaaggcagc aagctgcttc ctctgctgcc tgaaatacca gattcccaat ggcgaagatt 120
gagaaaaacg ctcccacgat ggaaaaaaag ccagaactgt ttaacatcat ggaagtagat 180
ggagtcccta cgttgatatt atcaaaagaa tggtgggaaa aagtctgtaa tttccaagcc 240
aagcctgatg atcttattct ggcaacttac ccaaagtcag gtacaacatg gatgcatgaa 300
attttagaca tgattctaaa tgatggtgat gtggagaaat gcaaaagagc ccagactcta 360
gatagacacg ctttccttga actgaaattt ccccataaag aaaaaccaga tttggagttc 420
gttcttgaaa tgtcctcacc acaactgata aaaacacatc tcccttcaca tctgattcca 480
ccatctatct ggaaagaaaa ctgcaagatt gtctatgtgg ccagaaatcc caaggattgc 540
ctggtgtcct actaccactt tcacaggatg gcttccttta tgcctgatcc tcagaactta 600
gaggaatttt atgagaaatt catgtccgga aaagttgttg gcaggtcctg gtttgaccat 660
gtgaaaggat ggtgggctgc aaaagacacg caccggatcc tctacctctt ctacgaggat 720
attaaaaaaa atccaaaaca tgagatccac aaggtgttgg aattcttgga gaaaactttg 780
tcaggtgatg ttataaacaa gattgtccac catacctcat ttgatgtaat gaaggataat 840
cccatggcca accatactgc ggtacctgct cacatattca atcactccat ctcaaaattt 900
atgaggaaag ggatgcctgg agactggaag aaccacttta ctgtggctat gaatgagaac 960
tttgataagc attatgaaaa gaagatggca gggtccacac tgaacttctg cctggagatc 1020
tgagaggaac aacaacaaac taggtgacag agactatgcc aactatttcg ccttttattc 1080
tgttgagcaa ggaactgtga ctgaatgtgg agcttatgag cttcagtcca tctcctatag 1140
tgtggctagt ttgctataat attaaaacat gatttaaaat atcaacaaac cagttactcc 1200
agcaaataaa ataagagaat tagagagcag aaaaaaaaaa 1240

<210> 34
<211> 2275
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1844394CB1

<400> 34
gcaattcagt gaggttaaag gactggatgc atttgttctg agcctgctca ctctagatgg 60
tgaatcaatc tacagcctga cctcgaagcc tatactactg ttattagcac gcattatcct 120
agtgaatgta agacataaac tgacagctat tcagagcttg ccatggtgga ctttgagatg 180
tgtgaatatt catcagcatt tgcttgagga acgctcacct ctgcttttta ctcttgccga 240
aaactgtatt gatcaagtga tgaaactaca gaatctgttt gtagatgatt caggtcgata 300
tttggctatt caattccatc tggaatgtgc atatgtgttt ttatattatt atgagtacag 360
aaaagcaaaa gatcagttgg atattgctaa ggacatcagc caattacaaa ttgatttgac 420
aggtgctttg ggaaaaagaa cacggttcca ggaaaattat gtggcacaac tgattctaga 480
tgtaagaagg gaaggggatg tcctttcaaa ttgtgaattc actccagcac ccactcctca 540
ggaacattta accaagaatc ttgagctcaa tgatgacacc attctgaatg acataaagtt 600
agcagattgt gaacagttcc agatgccgga tctgtgtgct gaagagatcg ctattattct 660
tggaatctgc actaattttc aaaagaataa cccagtgcac acattaactg aagtggagct 720
tctggcattt acatcatgtt tgctttcaca accaaagttc tgggccattc agacatcagc 780
cttgatcctc cggacaaaac ttgagaaagg aagtactcgc cgagtggaac gtgcaatgag 840
gcagacacag gctcttgcag accaatttga agataaaact acatctgtat tggaacgcct 900
gaagattttc tattgctgtc aagtaccacc tcactgggcc attcagcgcc aacttgcaag 960
tttgctcttt gagttgggat gtaccagttc agcccttcag atatttgaaa agctagaaat 1020
gtgggaagat gttgtcattt gttatgaaag agccgggcag cacggaaagg cagaagaaat 1080
ccttagacaa gagctggaga aaaaagaaac gcctagttta tactgcttgc ttggagatgt 1140
cctcggagac cattcttgct atgacaaggc ctgggagttg tcccggtacc gcagtgctcg 1200
tgctcagcgc tccaaagccc tccttcatct tcggaacaag gagtttcaag agtgtgtaga 1260
gtgcttcgaa cgctcggtta agattaatcc catgcagctc ggggtgtggt tttctctcgg 1320
ttgtgcctat ttggccttgg aagactatca aggttcagca aaggcatttc agcgctgtgt 1380
gactctagaa cccgataatg ctgaagcttg gaacaatttg tcaacttcct atatccgatt 1440
aaaacaaaaa gtaaaagctt ttagaacttt acaagaagct ctcaagtgta actatgaaca 1500
ctggcagatt tggaaaaact acatcctcac cagcactgac gttggggaat tttcagaagc 1560
cattaaagct tatcaccggc tcttggactt acgtgacaaa tacaaagatg ttcaggtcct 1620
taaaattcta gtcagggcag tgattgatgg gatgactgat cgaagtggag atgttgcaac 1680
tggcctcaaa ggaaagctgc aggagttatt tggcagagtg acttcaagag tgacaaatga 1740
tggagaaatc tggaggctgt atgcccacgt atatggaaat gggcagagtg aaaagcctga 1800
tgaaaatgaa aaggcattcc agtgcctctc aaaggcatac aagtgtgaca cccagtccaa 1860
ttgttgggag aaagatatta catcatttaa ggaagttgtt caaagagcct taggacttgc 1920
acatgtggcc ataaaatgca gtaaaaacaa atccagttcc caagaagctg tacaaatgct 1980

ttcttctgtt cgactcaatt tacggggctt gttatctaaa gcaaagcaac tttttacaga 2040
tgtggcaact ggagaaatgt ccagggaatt agctgatgac ataacagcta tggacacctt 2100
agtgacagag ctccaagacc taagcaacca gtttcgaaat cagtattgat tctgctggaa 2160
gcagattctg gaaaaggtgc tttcacctgc tggtaaaaga tacatctgta tatctgaaat 2220
gcaagatatt gatttttaaa ataaatttgt tttatgactt aaaaaaaaaa aaaaa 2275

<210> 35
<211> 1586
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2613056CB1

<400> 35
tctaaggcac agtatcattt tcagtactga caaggtgttt cattttatat ggttgtcata 60
ataaggcaaa ttcattttgt acgctttata ttttcaaacc cagcaagctc taaaagggac 120
ataaaataac ttagaaattg ggaaagacgg gcatgtgtat gatcatgata ttcatcccct 180
gccccagaac aaatgggagg aacacattgc ccaaaactca cgtctggagc tctttcaaca 240
tgtctccctg atgaccctgg acagcatcat gaagtgtgcc ttcagccacc agggcagcat 300
ccagttggac agtaccctgg actcatacct gaaagcagtg ttcaacctta gcaaaatctc 360
caaccagcgc atgaacaatt ttctacatca caacgacctg gttttcaaat tcagctctca 420
aggccaaatc ttttctaaat ttaaccaaga acttcatcag ttcacagaga aagtaatcca 480
ggaccggaag gagtctctta aggataagct aaaacaagat actactcaga aaaggcgctg 540
ggattttctg gacatacttt tgagtgccaa aagcgaaaac accaaagatt tctctgaagc 600
agatctccag gctgaagtga aaacgttcat gtttgcagga catgacacca catccagtgc 660
tatctcctgg atcctttact gcttggcaaa gtaccctgag catcagcaga gatgccgaga 720
tgaaatcagg gaactcctag gggatgggtc ttctattacc tgggaacacc tgagccagat 780
gccttacacc acgatgtgca tcaaggaatg cctccgcctc tacgcaccgg tagtaaacat 840
atcccggtta ctcgacaaac ccatcacctt tccagatgga cgctccttac ctgcaggaat 900
aactgtgttt atcaatattt gggctcttca ccacaacccc tatttctggg aagaccctca 960
ggtctttaac cccttgagat tctccaggga aaattctgaa aaaatacatc cctatgcctt 1020
cataccattc tcagctggat taaggaactg cattgggcag cattttgcca taattgagtg 1080
taaagtggca gtggcattaa ctctgctccg cttcaagctg gctccagacc actcaaggcc 1140
tccccagcct gttcgtcaag ttgtcctcaa gtccaagaat ggaatccatg tgtttgcaaa 1200
aaaagtttgc taattttaag tcctttcgta taagaattaa tgagacaatt ttcctaccaa 1260
aggaagaaca aaaggataaa tataatacaa aatatatgta tatggttgtt tgacaaatta 1320
tataacttag gatacttctg actggttttg acatccatta acagtaattt taatttcttt 1380
gctgtatctg gtgaaaccca caaaaacacc tgaaaaaact caagctgact tccactgcga 1440
agggaaatta ttggtttgtg taactagtgg tagagtggct ttcaagcata gtttgatcaa 1500
aactccactc agtatctgca ttacttttat ctctgcaaat atctgcatga tagctttatt 1560
ctcagttatc tttccccata ataaaa 1586

<210> 36
<211> 859
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5053617CB1

<400> 36
gtcgagtgcc tcccccaccc cccaccatgt gcttgagtgc acacccggcg ccaggccctg 60
atcctggcac ttcttgtgaa tcacaccgtg tcatacccat gacttccatt gcacagtggg 120
gaaactgagt ctagagaggt gaaataacat gtctaaagtc acaggaagtg aaaaagctga 180
ggacatggag ccagttgccc aatgacagga gagctgaaat gtcctcactg ctgggggtag 240
accgggcctc accagcttcc tggagagtca catgtttgtc tgcatcctca gggggctcgc 300
cggttctcca gcccggactg ctgccagagg cttcctggag gtggcagctt tcttcaaagg 360
caccatcccg gagcgcaagc ccctgatggg cgcagaaaat tcgggacaga ccacgtagag 420
gtgggctccc aagcaggtgc ggacggcacc aggccgccca aggcatcgct gccacctgag 480
ctccagccgc ccacaaactg ctgcatgagt ggctgcccca actgcgtgtg ggtggagtac 540
gcggacaggc tgctgcagca cttccaggac ggtggggagc gggccctggc tgccctggag 600

gagcacgtgg ctgatgagaa cctcaaggcc ttcctcagga tggagatccg gctgcacacc 660
aggtgcggag gctgagccat ccctgctgga ctccctaccg caggacggag tccaggacgc 720
agccgcagcc tccttccttc acacccccctc acagactcct tgtgtccaac gggaatagga 780
agaattagtt actgacttca cctgagaaaa aaataaattc tctatggtgg tttcaaaaaa 840
aaaaaaaaaa aaaaaaaaa 859

<210> 37
<211> 2302
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5483256CB1

<400> 37
gtttccgggg ctttcagtgg ccggaagtcg cggcgcctgt actgactcta ggaagggctg 60
gagtgttttg aatgggcgcc cgtaagagag gtgggcaagt acgtgttaca gacggccacg 120
ccgcccttta ggcggtcaag gtggggcgag cagacgttcg ccccccctgca gtcggccggg 180
tcactaccca agagcctttg gaggcggaag catggaacgg tctgcaaacg ttcccgagcg 240
ggcctctgcg gctctggcgg gcgtttcgaa cttgggcgcc gggcacacgc ccagtcccga 300
gagcgctgag ggttccctta gcgtcgccct caccccggcc aacccgcggg gcgccagagt 360
cctggccctt taaacgccgc gcgtgcctcg gcgtcttcgt ttcgcgcgcc cggccgcggc 420
gccggcggag cgaacatgga cccggctgcg cgggtggtgc gggcgctgtg gcctggtggg 480
tgcgccttgg cctggaggct gggaggccgc ccccagccgc tgctacccac gcagagccgg 540
gctggcttcg cgggggcggc gggcggcccg agccccgtgg ctgcagctcg taaggggagc 600
ccgcggctgc tgggagctgc ggcgctggcc ctgggggggag ccctggggct gtaccacacg 660
gcgcggtggc acctgcgcgc ccaggacctc cacgcagagc gctcagccgc gcagctctcc 720
ctgtccagcc gcctgcagct gaccctgtac cagtacaaga cgtgtccctt ctgcagcaag 780
gtccgagcct tcctcgactt ccatgccctg ccctaccagg tggtggaggt gaaccctgtg 840
cgcagggctg agatcaagtt ctcctcctac agaaaggtgc ccatcctggt ggcccaggaa 900
ggagaaagct cgcaacaact aaatgactcc tctgtcatca tcagcgccct caagacctac 960
ctggtgtcgg ggcagcccct ggaagagatc atcacctact acccagccat gaaggctgtg 1020
aacgagcagg gcaaggaggt gaccgagttc ggcaataagt actggctcat gctcaacgag 1080
aaggaggccc agcaagtgta tggtgggaag gaggccagga cggaggagat gaagtggcgg 1140
cagtgggcgg acgactggct ggtgcacctg atctccccca atgtgtaccg cacgcccacc 1200
gaggctctgg cgtcctttga ctacattgtc cgcgagggca agttcggagc cgtggagggt 1260
gccgtggcca agtacatggg tgcagcggcc atgtacctca tcagcaagcg actcaagagc 1320
aggcaccgcc tccaggacaa cgtgcgcgag gacctctatg aggctgctga caagtgggtg 1380
gctgctgtgg gcaaggaccg gcccttcatg gggggccaga agccgaatct cgctgatttg 1440
gcggtgtatg gcgtgctgcg tgtgatggag gggctggatg cattcgatga cctgatgcag 1500
cacacgcaca tccagccctg gtacctgcgg gtggagaggg ccatcaccga ggcctcccca 1560
gcgcactgaa tgtccccgcg cagagcagag ggaaggcagc ggaagacgcc agctgccagg 1620
gcctggggcc actgggccag cgcctggcga tactggttgg gggcaggatc attctgcccc 1680
ttgtccacgc acccccacca gccctctcgc ttctaacaca gggcacctgc tggggctcag 1740
ggatgttagg gacgagttcc agccctgcca ctgccctggg gcgaccccctc cctgtccctg 1800
cctccctgct ctgccgcccc tcttcctgga ccctcagtgg ctgtcccatg gctacatcct 1860
gtgggtgggg gccctcgaca ggacagcagg acggtttgtt ttcagtggaa tcccatccct 1920
gggttccccct ggttcccact cttcccaagc ctcccgggac tgggacatgt ttgcaataaa 1980
ggaaaggttt gtggcgcctg tcatggcagg catctcatgg agctccgtgt ggctgagtgc 2040
tgcgtggggc tggcggtcaa gggaggcatc aggcttgggc tgtgccagcc cttgtggtaa 2100
ctaaccgctg gcctggggct tcccaggtgt caggcacggt acggctccgc aggctttgtg 2160
tggcatcgtc cccaggatac cactcagggc acacagctgg gccgtggagc ccagcagcca 2220
gagtgcaggt cggggcaccc tacccacggt ggggctctgc agtggggtca ctcatcaagc 2280
ctcagtttct tcgtcatgtc cc 2302

<210> 38
<211> 1653
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5741354CB1

<400> 38
ctgggtctca gggctgctgt ggagttcgca cctccagctc gggccgatgt ggaagctttg 60
gagagctgaa gagggcgcgg cggcgctcgg cggcgcgctc ttcctgctgc tcttcgcgct 120
aggggtccgc cagctgctga agcagaggcg gccgatgggc ttccccccgg ggccgccggg 180
gctgccattt atcggcaaca tctattccct ggcagcctca tccgagcttc cccatgtcta 240
catgagaaag cagagccagg tgtacggaga gatcttcagt ttagatcttg gaggcatatc 300
aactgtggtt ctaaatggct atgatgtagt aaaggaatgc cttgttcatc aaagcgaaat 360
ttttgcagac agaccatgcc ttcctttatt catgaagatg acaaaaatgg gaggcttact 420
caattccaga tatggccgag gatgggttga tcacagacga ttagctgtaa acagttttcg 480
atattttgga tatggccaaa agtcttttga atctaaaatc ttggaagaaa ccaaattttt 540
caatgatgct attgaaacat acaaaggtag acctttttgac tttaaacagt taataacgaa 600
tgctgtttca aacataacca atctgatcat ttttggagaa cgattcactt atgaagacac 660
cgattttcag cacatgattg agttatttag tgaaaatgtg gaactagctg ccagtgcctc 720
agtcttcttg tataatgcct ttccatggat tggcatcctg ccttttggaa aacatcaaca 780
gctgtttaga aatgcagctg tagtctatga ttttctctcc agactcattg aaaaagcttc 840
agtcaacaga aagcctcagc tacctcagca ttttgttgat gcttatttag atgagatgga 900
tcaaggtaaa aatgacccat catctacttt ctccaaagaa aacctaattt tctcagtggg 960
tgaactcatc attgctggaa ctgaaactac aaccaatgtg ctacggtggg cgattctttt 1020
catggccctt tatcctaata ttcaaggaca agttcagaaa gagattgatt taattatggg 1080
ccctaatggg aagccttctt gggacgacaa atgcaaaatg ccttatactg aggcagtttt 1140
gcatgaagtt ttaagattct gtaatatagt tccattaggg attttccatg caacctctga 1200
agatgcagtt gtacgtggtt attccattcc taaaggcaca acagtaatta caaatcttta 1260
ttctgtacac tttgatgaaa agtactggag agacccagaa gtgttccatc ctgagcgatt 1320
tctggacagc agtggatatt ttgccaagaa ggaagctttg gttccttttt ccctaggaag 1380
aagacattgt cttggagaac acttggctcg gatggaaatg ttcttgtttt ttacagcatt 1440
gcttcagagg tttcatttgc attttccaca tgaactagtt ccagatctga agcccaggtt 1500
aggcatgaca ttgcagcccc aaccctacct catctgtgct gaaagacgct gaaactgcct 1560
gggatgtttt cgggaacaag aatgtatatt tgccttatcc ctgaacttgg tttaatcaaa 1620
tcaatgtgtg tattagaata aaagtcacag cat 1653

<210> 39
<211> 683
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5872615CB1

<400> 39
cgcgcgaccc cggactccac ggaggccgcg gcgagcaggc ggagctgcgg gtcgggacgc 60
tctgcgtggg gcggggcgca agggaggttt cgagcccgga aggtccggcg cccagagcta 120
acgggagtcc caggttaaac actttaagat gagaaaaatt gatctctgtc tgagctctga 180
agggtccgaa gtgattttag ctacatcaag tgatgaaaaa cacccacctg aaaatatcat 240
tgatgggaat ccagaaacgt tttggaccac cacaggaatg tttccccagg aattcattat 300
ttgtttccac aaacatgtaa ggattgaaag gcttgtaatc caaagttact ttgtacagac 360
cttgaagatt gaaaaaagca cgtctaaaga gccagttgat tttgagcaat ggattgaaaa 420
agatttggta cacacagagg ggcagcttca aaatgaagaa attgtggcac atgatggctc 480
cgctacttac ttgagattca ttattgtatc agcctttgat cattttgcat ctgtgcatag 540
cgtttctgca gaaggaacag tagtctcaaa tctttcctca taatgataac aaaatgctct 600
tgcatgattt tttaacaata tatttaaaca ggaagttgtc actgatatac tttattaaaa 660
ggatttttat caaaaaaaaa aaa 683

<210> 40
<211> 657
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2657543CB1

<400> 40
atgctatcca cttttgccag gcagaatgac atccctttc agctgcagac agtggagttg 60

gcttgggggg agcacctgaa gcctgagttc ctgaaggtga acccccctggg gaaggtgcct 120
gccctcagag atggcgactt cctactagca gagaggctgg agaaaagatc tctgacaccc 180
cctgcccaca gcatggtcat cgttttatac ctgagtcgaa agtaccagat acggggacac 240
tggtacccac ctgagctgca agcccgcacc tgcgtggatg agtacttggc gtggaagcat 300
gtcaccatcc agctgcctgc caccaatgtc tacctgtgca agcctgcaga tgctgcacag 360
ctggagcggc tgttggggag gctgacgcca gccctgcagc acctggatgg gggggtcctg 420
gtggccaggc ccttcctggc aatggagcag atctccctgg aagacttggt gctgacggag 480
gtgatgcagg tgaagctttc ctacccacct gccctcgggg ggactctggg catggggctg 540
agccccaacc ccagctgccc tgtcttccca gcccactgcc gttggctgcg acctcttcca 600
agactggccc tggctggcag tgtgacaggc ccatatgaag gctgcccttg gtactga 657

<210> 41
<211> 1122
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3041639CB1

<400> 41
tggatctgcg ggaatgtggg ctggagaggt cctgccgtgg taccagcctc cagcctgccc 60
ccaggactgc ccctgaccca ggcgcgcccg ctgctcggtg gcaggagggc cggcggagcg 120
ccatggcctg catcctgaag agaaagtctg tgattgctgt gagcttcata gcagcgttcc 180
ttttcctgct ggttgtgcgt cttgtaaatg aagtgaattt cccattgcta ctaaactgct 240
ttggacaacc tggtacaaag tggataccat tctcctacac atacaggcgg ccccttcgaa 300
ctcactatgg atacataaat gtgaagacac aagagccttt gcaactggac tgtgaccttt 360
gtgccatagt gtcaaactca ggtcagatgg ttggccagaa ggtgggaaat gagatagatc 420
gatcctcctg catttggaga atgaacaatg cccccaccaa aggttatgaa gaagatgtcg 480
gccgcatgac catgattcga gttgtgtccc ataccagcgt tcctcttttg ctaaaaaacc 540
ctgattattt tttcaaggaa gcgaatacta ctatttatgt tatttgggga cctttccgca 600
atatgaggaa agatggcaat ggcatcgttt acaacatgtt gaaaaagaca gttggtatct 660
atccgaatgc ccaaatatac gtgaccacag agaagcgcat gagttactgt gatggagttt 720
ttaagaagga aactgggaag gacagtacag agtgaccatg cagtgttgat tgatcgaaca 780
gcaaccacca catacatgtc ctgccccacc acaaaaggaa ggaaggaata aaagaaagaa 840
agaaagaaac aaacaaacaa acaaacaaaa ctaagcaaga caaaacaaat acccatgtca 900
gtggttcaaa gattaagatt gtggctttgt gtaaagttct ttccctttgt agacttgctg 960
cataattatt caggtatgat ggttacagtt tttaaaaagg aagggaaatt gtggtatgtg 1020
gtatgtaaat atttttaaat gttgtctctc tgttttgatc agtttttgtt ttattcaatt 1080
tgtctttatt aaatcttatc caagccaaaa aaaaaaaaaa ag 1122

<210> 42
<211> 2982
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3595451CB1

<400> 42
agccggtacc ggcgggcagg aggcgcccga ggatgtgctg ctggccgctg ctcctgctgt 60
gggggctgct ccccgggacg gcggcggggg gctcgggccg aacctatccg caccggaccc 120
tcctggactc ggagggcaag tactggctgg gctggagcca gcggggcagc cagatcgcct 180
tccgcctcca ggtgcgcact gcaggctacg tgggcttcgg cttctcgccc accggggcca 240
tggcgtccgc cgacatcgtc gtgggcgggg tggcccacgg gcggccctac ctccaggatt 300
attttacaaa tgcaaataga gagttgaaaa aagatgctca gcaagattac catctagaat 360
atgccatgga aaatagcaca cacacaataa ttgaatttac cagagagctg catacatgtg 420
acataaatga caagagtata acggatagca ctgtgagagt gatctgggcc taccaccatg 480
aagatgcagg agaagctggt cccaagtacc atgactccaa taggggcacc aagagtttgc 540
ggttattgaa tcctgagaaa actagtgtgc tatctacagc cttaccatac tttgatctgg 600
taaatcagga cgtccccatc ccaaacaaag atacaacata ttggtgccaa atgtttaaga 660
ttcctgtgtt ccaagaaaag catcatgtaa taaaggttga gccagtgata cagagaggcc 720
atgagagtct ggtgcaccac atcctgctct atcagtgcag caacaacttt aacgacagcg 780

ttctggagtc cggccacgag tgctatcacc ccaacatgcc cgatgcattc ctcacctgtg 840
aaactgtgat ttttgcctgg gctattggtg gagagggctt ttcttatcca cctcatgttg 900
gattatccct tggcactcca ttagatccgc attatgtgct cctagaagtc cattatgata 960
atcccactta tgaggaaggc ttaatagata attctggact gaggttattt tacacaatgg 1020
atataaggaa atatgatgct ggggtgattg aggctggcct ctgggtgagc ctcttccata 1080
ccatccctcc agggatgcct gagttccagt ctgagggtca ctgcactttg gagtgcctgg 1140
aagaggctct ggaagccgaa aagccaagtg gaattcatgt gtttgctgtt cttctccatg 1200
ctcacctggc tggcagaggc atcaggctgc gtcattttcg aaaagggaag gaaatgaaat 1260
tacttgccta tgatgatgat tttgacttca atttccagga gtttcagtat ctaaaggaag 1320
aacaaacaat cttaccagga gataacctaa ttactgagtg tcgctacaac acgaaagata 1380
gagctgagat gacttgggga ggactaagca ccaggagtga aatgtgtctc tcataccttc 1440
tttattaccc aagaattaat cttactcgat gtgcaagtat tccagacatt atggaacaac 1500
ttcagttcat tggggttaag gagatctaca gaccagtcac gacctggcct ttcattatca 1560
aaagtcccaa gcaatataaa aacctttctt tcatggatgc tatgaataag tttaaatgga 1620
ctaaaaagga aggtctctcc ttcaacaagc tggtcctcag cctgccagtg aatgtgagat 1680
gttccaagac agacaatgct gagtggtcga ttcaaggaat gacagcatta cctccagata 1740
tagaaagacc ctataaagca gaacctttgg tgtgtggcac gtcttcttcc tcttccctgc 1800
acagagattt ctccatcaac ttgcttgttt gccttctgct actcagctgc acgctgagca 1860
ccaagagctt gtgatcaaaa ttctgttgga cttgacaatg ttttctatga tctgaacctg 1920
tcatttgaag tacaggttaa agactgtgtc cactttgggc atgaagagtg tggagacttt 1980
tcttccccat tttccctccc tccttttttcc tttccatgtt acatgagaga catcaatcag 2040
gttctcttct ctttcttaga aatatctgat gttatatata catggtcaat aaaataaaac 2100
tggcctgact taagataacc attttaaaaa attgggctgt catgtgggaa taaaagaatt 2160
ctttctttcc tactacattc tgttttattt aaatactcat tgttgctatt tcacttttg 2220
acttgacttt tatatttctt taaaaaattc cttcctttta aaaaatataa aagggactac 2280
tgttcattcc agttttcttc ttctttgttg ttcttctagt gtgacttttc aagtgtaaca 2340
gccattcttc ctgactttaa tattgtccag ttctggtctt ttctgtgaat taccactggg 2400
ccccttacct caatgctttt tgttgatgcc cactctggtt cccttgttta tctgagtctg 2460
ttggtacccc aaatgacccc acacccatct taaagtactt tttttcacct tccctgttta 2520
gtactggcca gatgagtttt ttctagagct ctgtcactat ctgaaaagaa agaggctatg 2580
ggaaacatag aaatggtatg tattaataac tgatcatagg ctgaggagaa aaaatgtagc 2640
tggctgcaaa cccagtgctg tgaggtgact tatatgaggt tccagatcaa agacaggccg 2700
tgtgagccag tccaggaggg tgtaagttct gaatggttcc ttgctgactt tgggtgacac 2760
atgtaccaca tactggctca gtttaagtca tggttctatt gtagatttat ttttatatta 2820
gttaataaat gactttaaat tgtcaccaat tgaaaatctt gtcactcttt tggttttctt 2880
tatatagctc agccaaatct tgttttatgt cctgtcctca tctcttaagc taaatctgtt 2940
tggatcatat taataaacta aatgaaatta aaaaaaaaaa aa 2982

<210> 43
<211> 3517
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4169101CB1

<400> 43
tggccactat tacggcgcag tgtgctggaa aggcggggct caggctcctt gcagattcct 60
aaccagcata atgctggagc cgggagccac caacctgcag ttttcagaat ggccgtgttg 120
gacactgatt tggatcacat tcttccatct tctgttcttc ctccattctg ggctaagtta 180
gtagtgggat cggttgccat tgtgtgtttt gcacgcagct atgatggaga ctttgtcttt 240
gatgactcag aagctattgt taacaataag gacctccaag cagaaacgcc cctgggggac 300
ctgtggcatc atgacttctg gggcagtaga ctgagcagca acaccagcca caagtcctac 360
cggcctctca ccgtcctgac tttcaggatt aactactacc tctcgggagg cttccacccc 420
gtgggctttc acgtggtcaa catcctcctg cacagtggca tctctgtcct catggtggac 480
gtcttctcgg ttctgtttgg cggcctgcag tacaccagta aaggccggag gctgcacctc 540
gcccccaggg cgtccctgct ggccgcgctg ctgtttgctg tccatcctgt gcacaccgag 600
tgtgttgctg gtgttgtcgg ccgtgcagac ctcctgtgtg ccctgttctt cttgttatct 660
ttccttggct actgtaaagc atttagagaa agtaacaagg agggagcgca ttcttccacc 720
ttctgggtgc tgctgagtat ctttctggga gcagtggcca tgctgtgcaa agagcaaggg 780
atcactgtgc tgggtttaaa tgcggtattt gacatcttgg tgataggcaa attcaatgtt 840
ctggaaattg tccagaaggt actacataag gacaagtcat tagagaatct cggcatgctc 900
aggaacgggg gcctcctctt cagaatgacc ctgctcacct ctggaggggc tgggatgctc 960

tacgtgcgct ggaggatcat gggcacgggc ccgccggcct tcaccgaggt ggacaacccg 1020
gcctcctttg ctgacagcat gctggtgagg gccgtaaact acaattacta ctattcattg 1080
aatgcctggc tgctgctgtg tccctggtgg ctgtgttttg attggtcaat gggctgcatc 1140
cccctcatta agtccatcag cgactggagg gtaattgcac ttgcagcact ctggttctgc 1200
ctaattggcc tgatatgcca agccctgtgc tctgaagacg gccacaagag aaggatcctt 1260
actctgggcc tgggatttct cgttatccca tttctccccg cgagtaacct gttcttccga 1320
gtgggcttcg tggtcgcgga gcgtgtcctc tacctcccca gcattgggta ctgtgtgctg 1380
ctgacttttg gattcggagc cctgagcaaa cataccaaga aaaagaaact cattgccgct 1440
gtcgtgctgg gaatcttatt catcaacacg ctgagatgtg tgctgcgcag cggcgagtgg 1500
cggagtgagg aacagctttt cagaagtgct ctgtctgtgt gtcccctcaa tgctaaggtt 1560
cactacaaca ttggcaaaaa cctggctgat aaaggcaacc agacagctgc catcagatac 1620
taccgggaag ctgtaagatt aaatcccaag tatgttcatg ccatgaataa tcttggaaat 1680
atcttaaaag aaaggaatga gctacaggaa gctgaggagc tgctgtcttt ggctgttcaa 1740
atacagccag actttgccgc tgcgtggatg aatctaggca tagtgcagaa tagcctgaaa 1800
cggtttgaag cagcagagca aagttaccgg acagcaatta aacacagaag gaaataccca 1860
gactgttact acaacctcgg gcgtctgtat gcagatctca atcgccacgt ggatgccttg 1920
aatgcgtgga gaaatgccac cgtgctgaaa ccagagcaca gcctggcctg gaacaacatg 1980
attatactcc tcgacaatac aggtaattta gcccaagctg aagcagttgg aagagaggca 2040
ctggaattaa tacctaatga tcactctctc atgttctcgt tggcaaacgt gctggggaaa 2100
tcccagaaat acaaggaatc tgaagcttta ttcctcaagg caattaaagc aaatccaaat 2160
gctgcaagtt accatggtaa tttggctgtg ctttatcatc gttggggaca tctagacttg 2220
gccaagaaac actatgaaat ctccttgcag cttgacccca cggcatcagg aactaaggag 2280
aattacggtc tgctgagaag aaagctagaa ctaatgcaaa agaaagctgt ctgatcctgt 2340
ttccttcatg ttttgagttt gagtgtgtgt gtgcatgagg catatcatta atagtatgtg 2400
gttacattta accatttaaa agtcttagac atgttatttt actgattttt ttctatgaaa 2460
acaaagacat gcaaaaagat tatagcacca gcaatatact cttgaatgcg tgatatgatt 2520
tttcattgaa attgtatttt ttcagacaac tcaaatgtaa ttctaaaatt ccaaaaatgt 2580
ctttttaat taaacagaaa aagagaaaaa attatcttga gcaactttta gtagaattga 2640
gcttacattt gggatctgag ccttgtcgtg tatggactag cactattaaa cttcaattat 2700
gaccaagaaa ggatacactg gcccctacaa tttgtataaa tattgaacat gtctatatat 2760
tagcattttt atttaatgac aaagcaaatt aagttttttt atctcttttt tttaaaacaa 2820
catactgtga actttgtaag gaaatattta tttgtatttt tatgttttga atagggcaaa 2880
taatcgaatg aggaatggaa gttttaacat agtatatcta tatgcttttc cccataggaa 2940
gaaattgact cttgcagttt ttggatgctc tgacttgtgc aatttcaata cacaggagat 3000
tatgtaatgt aatatttttc ataagcggtt actatcaatt gaaagttcaa gccatgcttt 3060
aggcaagagc aggcagcctc acatctttat ttttgttaca tccaaggtga agagggcaac 3120
acatctgtgt aagctgcttt ttagtgtgtt tatctgaagg ccgttttcca ttttgcttaa 3180
tgtaactaca gacattatcc agaaaatgca aaattttcta tcaaatggag ccacattcgg 3240
ggaattcgtg gtatttttaa gaattgagtt gttcctgctg ttttttattt gatccaaaca 3300
atgttttgtt ttgttcttct ctgtatgctg ttgacctaat gatttatgca atctctgtaa 3360
tttcttatgc agtaaaatta ctacacaaac tagcatgaaa atgtcatatt gccttcttaa 3420
tcaattattt tcaagtagtg aactttgtat cctcctttac cttaaaatga aatcaaactg 3480
accaaatcat catttatgtg gcttctgtgt gacttgg 3517

<210> 44
<211> 2339
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2925182CB1

<400> 44
ggcagccgcg ggagcacggc gacgccagcg gggtgaaggg aaaaggccga ggcatcagcg 60
tgtgaagacc gcaaagacga tcccgagtac agttgtgaac agcattgctg ctaggctcct 120
cctgcagatc atctgaaatg aacctctctt attgattttt attggcctag agccaggagt 180
actgcattca gttgactttc agggtaaaaa gaaaacagtc ctggttgttg tcatcataaa 240
catatggacc agtgtgatgg tgaaatgaga tgaggctccg caatggaact gtagccactg 300
ctttagcatt tatcacttcc ttccttactt tgtcttggta tactacatgg caaaatggga 360
aagaaaaact gattgcttat caacgagaat tccttgcttt gaaagaacgt cttcgaatag 420
ctgaacacag aatctcacag cgctcttctg aattaaatac gattgtgcaa cagttcaagc 480
gtgtaggagc agaaacaaat ggaagtaagg atgcgttgaa taagttttca gataataccc 540
taaagctgtt aaaggagtta acaagcaaaa aatctcttca agtgccaagt atttattatc 600

atttgcctca tttattgaaa aatgaaggaa gtcttcaacc tgctgtacag attggcaacg 660
gaagaacagg agtttcaata gtcatgggca ttcccacagt gaagagagaa gttaaatctt 720
acctcataga aactcttcat tcccttattg ataacctgta tcctgaagag aagttggact 780
gtgttatagt agtcttcata ggagagacag atattgatta tgtacatggt gttgtagcca 840
acctggagaa agaattttct aaagaaatca gttctggctt ggtggaagtc atatcacccc 900
ctgaaagcta ttatcctgac ttgacaaacc taaaggagac atttggagac tccaaagaaa 960
gagtaagatg gagaacaaag caaaacctag attactgttt tctaatgatg tatgctcaag 1020
aaaagggcat atattacatt cagcttgaag atgatattat tgtcaaacaa aattatttta 1080
ataccataaa aaattttgca cttcaacttt cttctgagga atggatgatt ctagagtttt 1140
cccagctggg cttcattggt aaaatgtttc aagcgccgga tcttactctg attgtagaat 1200
tcatattcat gttttacaag gagaaaccca ttgattggct cctggaccat attctctggg 1260
tgaaagtctg caaccctgaa aaagatgcaa aacattgtga tagacagaaa gcaaatctgc 1320
gaattcgctt cagaccttcc cttttccaac atgttggtct gcactcatca ctatcaggaa 1380
aaatccaaaa actcacggat aaagattata tgaaaccatt acttcttaaa atccatgtaa 1440
acccacctgc ggaggtatct acttccttga aggtctacca agggcatacg ctggagaaaa 1500
cttacatggg agaggatttc ttctgggcta tcacaccgat agctggagac tacatcttgt 1560
ttaaatttga taaaccagtc aatgtagaaa gttatttgtt ccatagcggc aaccaagaac 1620
atcctggaga tattctgcta aacacaactg tggaagtttt gccttttaag agtgaaggtt 1680
tggaaataag caaagaaacc aaagacaaac gattagaaga tggctatttc agaataggaa 1740
aatttgagaa tggtgttgca gaaggaatgg tggatccaag tctcaatccc atttcagcct 1800
ttcgactttc agttattcag aattctgctg tttgggccat tcttaatgag attcatatta 1860
aaaaagccac caactgatca tctgagaaac caacacattt tttcctgtga atttgttaat 1920
taaagatagt taagcatgta tctttttttt atttctactt gaacactacc tcttgtgaag 1980
tctactgtag ataagacgat tgtcgtttcc acttggaaag tgaatctccc ataataattg 2040
tatttgtttg aaactaagct gtcctcagat tttaacttga ctcaaacatt tttcaattat 2100
gacagcctgt taatatgact tgtactattt tgggtattat actaataaca taagagttgt 2160
acatattgtt acattcttta aatttgagaa aaactaatgt tacatacatt ttatgaaggg 2220
gggtactttt gagattcact tattttacta ttatagaccc tcttttatag attatcaggg 2280
attatatata taaatatata aatatacata aaaatgttat ggattaattt attagaaca 2339

<210> 45
<211> 1955
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3271838CB1

<400> 45
gccccagaag ccccacgacg atggcggcaa tggcggtggc gctgcgggga ttaggagggc 60
gcttccggtg gcggacgcag gccgtggcgg gcggggtgcg gggcgcggcg cggggcgcag 120
cagcaggtca gcgggactat gatctcctgg tggtcggcgg gggatctggt ggcctggctt 180
gtgccaagga ggccgctcag ctgggaagga aggtgtccgt ggtggactac gtggaacctt 240
ctccccaagg cacccggtgg ggccttggcg gcacctgcgt caacgtgggc tgcatcccca 300
agaagctgat gcaccaggcg gcactgctgg gaggcctgat ccaagatgcc cccaactatg 360
gctgggaggt ggcccagccc gtgccgcatg actggaggaa gatggcagaa gctgttcaaa 420
atcacgtgaa atccttgaac tggggccacc gtgtccagct tcaggacaga aaagtcaagt 480
actttaacat caaagccagc tttgttgacg agcacacggt ttgcggcgtt gccaaaggtg 540
ggaaagagat tctgctgtca gccgatcaca tcatcattgc tactggaggg cggccgagat 600
accccacgca catcgaaggt gccttggaat atggaatcac aagtgatgac atcttctggc 660
tgaaggaatc ccctggaaaa acgttggtgg tcggggccag ctatgtggcc ctggagtgtg 720
ctggcttcct caccgggatt gggctggaca ccaccatcat gatgcgcagc atccccctcc 780
gcggcttcga ccagcaaatg tcctccatgg tcatagagca catggcatct catggcaccc 840
ggttcctgag gggctgtgcc ccctcgcggg tcaggaggct ccctgatggc cagctgcagg 900
tcacctggga ggaccgcacc accggcaagg aggacacggg caccttttgac accgtcctgt 960
gggccatagg tcgagtccca gacaccagaa gtctgaattt ggagaaggct ggggtagata 1020
ctagccccga cactcagaag atcctggtgg actcccggga agccacctct gtgccccaca 1080
tctacgccat tggtgacgtg gtggaggggc ggcctgagct gacacccaca gcgatcatgg 1140
ccgggaggct cctggtgcag cggctcttcg gcgggtcctc agatctgatg gactacgaca 1200
atgttcccac gaccgtcttc accccgctgg agtatggctg tgtggggctg tccgaggagg 1260
aggcagtggc tcgccacggg caggagcatg ttgaggtcta tcacgcccat tataaaccac 1320
tggagttcac ggtggctgga cgagatgcat cccagtgtta tgtaaagatg gtgtgcctga 1380
gggagccccc acagctggtg ctgggcctgc atttccttgg ccccaacgca ggcgaagtta 1440

ctcaaggatt tgctctgggg atcaagtgtg gggcttccta tgcgcaggtg atgcggaccg 1500
tgggtatcca tcccacatgc tctgaggagg tagtcaagct gcgcatctcc aagcgctcag 1560
gcctggaccc cacggtgaca ggctgctgag ggtaagcgcc atccctgcag gccagggcac 1620
acggtgcgcc cgccgccagc tcctcggagg ccagacccag gatggctgca ggccaggttt 1680
gggggcctc aaccctctcc tggagcgcct gtgagatggt cagcgtggag cgcaagtgct 1740
ggacaggtgg cccgtgtgcc ccacagggat ggctcagggg actgtccacc tcacccctgc 1800
acctctcagc ctctgccgcc gggcaccccc cccdaggctc ctggtgccag atgatgacga 1860
cctgggtgga aacctaccct gtgggcaccc atgtccgagc cccctggcat ttctgcaatg 1920
caaataaaga gggtactttt tctgaaaata aaaaa 1955

<210> 46
<211> 2065
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3292871CB1

<400> 46
ctaggcccta cttcgcagtt cttgtgcacg ctatgaaaaa taaaacctgc gtgctcgtct 60
gtgtgagtgt gtttggtggg gagagggggc aggtgactgt accccgggtt ggggtccgcc 120
gcccctccct cgcgggccct ctgcagaagt gcaccctgag agagacccgg gtgtggctcc 180
cgcagggttc tggcttccag tcgtcgcgga gggagaagta tggcaacgtg ttcaagacgc 240
atttgttggg gcggccgctg atacgcgtga ccggcgcgga gaacgtgcgc aagatcctca 300
tgggcgagca ccacctcgtg agcaccgagt ggcctcgcag cacccgcatg ttgctgggcc 360
ccaacacggt gtccaattcc attggcgaca tccaccgcaa caagcgcaag gtcttctcca 420
agatcttcag ccacgaggcc ctggagagtt acctgcccaa gatccagctg gtgatccagg 480
acacactgcg cgcctggagc agccaccccg aggccatcaa cgtgtaccag gaggcgcaga 540
agctgacctt ccgcatggcc atccgggtgc tgctgggctt cagcatccct gaggaggacc 600
ttgggcacct ctttgaggtc taccagcagt ttgtggacaa tgtcttctcc ctgcctgtcg 660
acctgccctt cagtggctac cggcggggca ttcaggctcg gcagatcctg cagaaggggc 720
tggagaaggc catccgggag aagctgcagt gcacacaggg caaggactac ttggacgccc 780
tggacctcct cattgagagc agcaaggagc acgggaagga gatgaccatg caggagctga 840
aggacgggac cctggagctg atctttgcgg cctatgccac cacggccagc gccagcacct 900
cactcatcat gcagctgctg aagcacccca ctgtgctgga gaagctgcgg gatgagctgc 960
gggctcatgg catcctgcac agtggcggct gcccctgcga gggcacactg cgcctggaca 1020
cgctcagtgg gctgcgctac ctggactgcg tcatcaagga ggtcatgcgc ctgttcacgc 1080
ccatttccgg cggctaccgc actgtgctgc agaccttcga gcttgatggt ttccagatcc 1140
ccaaaggctg gagtgtcatg tatagcatcc gggacaccca tgacacagcg cccgtgttca 1200
aagacgtgaa cgtgttcgac cccgatcgct tcagccaggc gcggagcgag gacaaggatg 1260
gccgcttcca ttacctcccg ttcggtggcg gtgtccggac ctgcctgggc aagcacctgg 1320
ccaagctgtt cctgaaggtg ctggcggtgg agctggctag caccagccgc tttgagctgg 1380
ccacacggac cttcccccgc atcaccttgg tccccgtcct gcaccccgtg gatggcctca 1440
gcgtcaagtt ctttggcctg gactccaacc agaacgagat cctgccggag acggaggcca 1500
tgctgagcgc cacagtctaa cccaagaccc acccgcctca gcccagccca ggcagcgggg 1560
tggtgcttgt gggaggtaga aacctgtgtg tgggaggggg ccggaacggg gagggcgagt 1620
ggcccccata cttgccctcc cttgctcccc cttcctggca aaccctaccc aaagccagtg 1680
ggccccattc ctagggctgg gctccccttc tggctccagc ttccctccag ccactcccca 1740
tttaccatca gctcagcccc tgggaagggc gtggcagggg ctctgcatgc ccgtgacagt 1800
gttaggtgtc agcgcgtgct acagtgtttt tgtgatgttc tgaactgctc ccttccctcc 1860
gttcctttcg gacccttta gctggggttg ggggacggga agagccgtgc ccccttgggc 1920
gcactcttca gcgtctcctc ctcctgcgcc cccactgcgt ctgcccagga acagcatcct 1980
gggtagcaga acaggagtca accttggcgg ggcgggggct gcgtccaacc tggagattgc 2040
ccttccctat gccacggttc ccacc 2065

<210> 47
<211> 866
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4109179CB1

<400> 47
ttcaaaagag gtttctggtc actcctaatc atcgcagcat aactctgctt tttaagctat 60
tgttttctgc atttgtaggg gcacgataca actgcagctg caataaactg gtccttatac 120
ctgttgggtt ctaacccaga agtccagaaa aaagtggatc atgaattgga tgacgtgttt 180
gggaagtctg accgtcccgc tacagtagaa gacctgaaga aacttcggta tctggaatgt 240
gttattaagg agacccttcg ccttttttcct tctgttcctt tatttgcccg tagtgttagt 300
gaagattgtg aagtggcagg ttacagagtt ctaaaaggca ctgaagccgt catcattccc 360
tatgcattgc acagagatcc gagatacttc cccaaccccg aggagttcca gcctgagcgg 420
ttcttccccg agaatgcaca agggcgccat ccatatgcct acgtgccctt ctctgctggc 480
cccaggaact gtataggtca aaagtttgct gtgatggaag aaaagaccat tctttcgtgc 540
atcctgaggc acttttggat agaatccaac cagaaaagag aagagcttgg tctagaagga 600
cagttgattc ttcgtccaag taatggcatc tggatcaagt tgaagaggag aaatgcagat 660
gaacgctaac tatattattg ggttgtgcct ttatcatgag aaaggtcttt attttaagag 720
atccttgtca tttacaattt acagatcatg agttcaatat gcttgaatcc cctagaccta 780
atttttcctt gatcccactg atcttgacat caagtctaac aaagaaaaag ttttgagttt 840
tgtattttct ttttttcttt ttcttt 866

<210> 48
<211> 1593
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4780365CB1

<400> 48
ttacggcgca gtgtgctgga cagcggtctc ccagggaagg gggtgctgag tggaaggagg 60
tcaatgggaa gccggggtgg ctctcagagt cggcaggagc agtcgggctg atgagctggg 120
aggagcagac cgcctccctc ttctctgagt gggaggaggg ccagatctgg actgggtttg 180
gagatgctca ggtggggctc agagcatcac ctgtggggca gagggaccat cttggcagat 240
gaaggcccgt cgcagggtgt gatgcctgaa ttacaaggcg ggacaggtaa agtggggcag 300
gtgagagaag gagggtgagt gatgtgattt ttctactcct gttttccagg aaaaccaaaa 360
tgccacgcac ttcgacctat gatccttttc ctaataatgc ttgtcttggt cttgtttggt 420
tacggggtcc taagccccag aagtctaatg ccaggaagcc tggaacgggg gttctgcatg 480
gctgttaggg aacctgacca tctgcagcgc gtctcgttgc caaggatggt ctacccccag 540
ccaaaggtgc tgacaccgtg taggaaggat gtcctcgtgg tgaccccttg gctggctccc 600
attgtctggg agggcacatt caacatcgac atcctcaacg agcagttcag gctccagaac 660
accaccattg ggttaactgt gtttgccatc aagaaatacg tggctttcct gaagctgttc 720
ctggagacgg cggagaagca cttcatggtg ggccaccgtg tccactacta tgtcttcacc 780
gaccagccgg ccgcggtgcc ccgcgtgacg ctggggaccg gtcggcagct gtcagtgctg 840
gaggtgcgcg cctacaagcg ctggcaggac gtgtccatgc gccgcatgga gatgatcagt 900
gacttctgcg agcggcgctt cctcagcgag gtggattacc tggtgtgcgt ggacgtggac 960
atggagttcc gcgaccacgt gggcgtggag atcctgactc cgctgttcgg caccctgcac 1020
cccggcttct acggaagcag ccgggaggcc ttcacctacg agcgccggcc ccagtcccag 1080
gcctacatcc ccaaggacga gggcgatttc tactacctgg gggggttctt cgggggggtcg 1140
gtgcaagagg tgcagcggct caccagggcc tgccaccagg ccatgatggt cgaccaggcc 1200
aacggcatcg aggccgtgtg gcacgacgag agccacctga acaagtacct gctgcgccac 1260
aaacccacca aggtgctctc ccccgagtac ttgtgggacc agcagctgct gggctggccc 1320
gccgtcctga ggaagctgag gttcactgcg gtgcccaaga accaccaggc ggtccggaac 1380
ccgtgagcgg ctgccagggg ctctgggagg gctgccggca gccccgtccc cctcccgccc 1440
ttggttttag cagaacgggt aaactctgtt tcctttgtcc gtcctgttgt gagtaactga 1500
agcctaggcc ccgtccccac ctcaaatcac acacaccccc tccccaccac agagacacca 1560
ttacatacac agacacacac agaaagacac aca 1593

Table 1

AATGATAGGTTTTTTTTTATATAGTTGGAGGGCAGGGATAGTGTTTAAAGGTTTATGATG
TATAAGCACGCTTAGACACACCAGGAGAAAGTATCAGATCTCATTACACATGGTTGTGAG
CCACCATGTGGTTGCTGGGATTTGAACTCGGGACCTCTGGAAGAGCAGTCAGTGCTCTTA
TCCACTGAGCCATCTCTATAGCACTTTAAAAAAAATATTAAGGCATCATACAGCNTTTCT
GCCTCTATGTATGCCCTGCA

Sequence 1104

CGGGCAGGTCTTTGTCTCTCAAACTTTTATACTGTCATAGTTCATTTTTCTTGTACAGAT
GCTCTTTCAACTTTTTTTGAGAGTATATACATAACAATTGCAGTTGAAAGTAGGGTGGGG
TTTAGTCTCCTGGTCACAGTTCAATAAGTACGGACGCGTGGGTCGACTCAAGCTAGCTTG
TACACCTN

Sequence 1105

TCGACTNCTATAGGGCGAATTGGAGCTCCCCGCGGTGGCGGCCGAGGTACACTTATAGTT
GAGAGCCAAGTCTCCCTTATCATTGGTGAATGAGGATGAGCTACTGAAAACAAAAAGAGG
GTCTTCTACTCAGCCTCTACCCCTAATATTTATATCAGAAGCAGAGATTAACTGTCCTTA
CTCATTCACACGTTAATGGAAGAGAAGGAAGTTTCCTAGAAAAATCCTCCCGCTCCACCC
TGCAAACTTTATGCTTTTCTGTTACATAATCAGGCAGGGGCAAGACCTAAACTATTTTGA
ATTGGTGGTGTTGAGGCTAAATTCTCTGCTATTGACAGAATTGAGAATGTGATCAATTTC
AGAGTAGCCATGTTACAAATTTTGTCCCAATTTCAATGGGGGAGGAATTATAACCAAATC
AAGTAGTGGTTGGGAAGA

Sequence 1106

CCGGGCAGGTACGCGGGGAGTTCTCTCAGGCTCTCCAGAGCTCAGGACCTCTGAGAAGAA
TGGAGCCCTCCTGGCTTCAGGAACTCATGGCTCACCCCTTCTTGCTGCTGATCCTCCTCT
GCATGTCTCTGCTGCTGTTTCAGGTAATCAGGTTGTACCT

Sequence 1107

TCACTACTTAGGGCGAATTGGAGCTCCCCGCGGTGGCGGCCGAGGTACGCGGGGGTAGAT
GGAAGGAAGAACTCGTGTGCTTAGACCTGACGCTGGGAGGAGATGCTGCCACCTAGGTTA
CTTGTAGGACCCTATACGGCAACCTCCTTTGCCAGGAACTATTTATAAACATCCTGCAGG
AAAATGCAGTAAAGTAGAAGAGACAGGGATATCCCAGAAGGTTATGCAAAACATCAAGAG
AAGATGAGAGGAGTCTATATGTCAGAATACACATTTCCCACCTTGCCCAACAGTAGAAAA
ACATAAGAAGAGAAAAACATTAAAAAATGACAAGGAAGTTAATGGAAGTCAGCAATGTGA
TGGTGTTTGGGAGGTGGAGCCCTTCAGAAAGGTAATTAATGCCCTTGTAAGAAAGAGGCC
AGNAGAGCTTGCCGCACC

Sequence 1108

GGGCAGGTACTTGTCCTGATGGGTCATCCAGAAGAGTTTCATGGAGATCAGTTCACAACA
TTGGACAGACTATGAAACAAAGTCTGTAATATCATTTATACTCAGTATTGGTAGGCTTTT
TCCTTAATACCCCCACTCCAACAAAACAGACCTGCCTTCACTTCTCTTGAAAATATTATT
ACCATCAGTAACTGCTTGTCTACTATGTCTCCAAGGCAAGCCACAGGCTATAAGCCAATT
CTGCAAAAGGGAGCAGCAAAAGAAATGCTTAGTGCAATTCCTCTATTTAAAAGTTTGCTT
GAGTTTTGATTATTTTTGTTATTTTG

Sequence 1109

AGGTCACGCGTCCGACCTAGCTTGAGTCGACCCACGCGTCCGATTTCTGGTTTCTTTGGA
GACTCGGTTGTGCCATGTGATGCTTTTCTGGAAACTGTCTTATGAGAGGAAGCTTTTGCT
GAAGCAGACACATGGGAGAATGTTTTGTTAAGAACAGACATGTAGTATTTTTCTAGAGGC
AGCTTGAATAAAGAGCATGTGGTATTTTGCTAGAGGGGACACTTGAGAGAACATGTGACA
TTTGGAAAGGATATATGTATAACCTAAAAGGACAGTGGCCAATGCTGTGTGGTATTTGGG
TGTGCCTTACCATTCTTTGCTGGC

Sequence 1110

TCGTGATCTAGATCCCTATAGTGAGCGGACGCGTGGGTCGACTCAAGCTAGGTCGGACGC
GTGGGTCGACTCAAGCTAGGTCGGACGCGTGACCTTCCCCAAAATTCTTCTTTTTTCAGA

Table 3

TATTTATAAACATCCTGCAGGAAAATGAGTCTATATGTCAGAATACACAT
TTCCCACCTTGCCCAACAGTAGAAAAACATAAGAAGAGAAAAACATTAAA
AAATGACAAGGAAGTTAATGGAAGTCAGCAATGTGATGGTGTTTGGAGGT
GGAGCCTTCAGAAGGTAATTAATGCCCTTGTAAGAAGAGGCCAGAGAGCT
TGCGCACCTTCTTCCTGCCATGTGAGGAGCCAAGAAGCCGGCTGTCTGCA
ACCTGCAAGAGGACCCTCACTAGAAGCTAGCCATACTGGCATCCTCATCT
TGGCTTTCCAACTTCCAGAACTGTGAGAAGTATATGTTTGTGGTTTAGTC
AATGGTCTATGGTAATTTTTTTATAGCAGTCCCAGCCAAGACAGTGCCTC
ATTTACTACATACCAATTATATTATTATATAGGCCTCTTTTCAGAAACCC
ATGTTCAAATAAGAGAATAGATACTGAAACACATAACACCTTCACTAGTT
TTTAGTATACAAATATTGAGAAATAGGTTGTTAATAACTATCTCAATCAA
GAAATGCCGAATCATGGTGGTTCTAATTTTTTATATATAAATTGACAAAA
TGGAGAAACTTAACACCATTCTAGATTTTAACTGCCCAAAGANAAANNAN
AAANNNAAGGGGGGGANGGACCTCGGCCGTTTTTA
>Sequence 3369
CCGGGCAGGTACTTTACTCACCCTTCCTCTGACAGAAAAGGATGAAGTCA
AGGGCCTGGTAGAGGCACCACTAAGAAAGGCGTCTGAGAGGACCAAAGAG
AGTGACCAGCAAGCANTTTTTTGCAAGGCTGAGGAGCTGACAGCTTCATG
AAAGGCTGGACCACCCAGTGGTGAGAAGCATCATCTGGGTTACCTTGTGC
TGCCATAAAACACACCACAGACTTGGTGACTTAAACCACAGATATTTATC
TTCTCACAATCCTGGAGGCTGGAAGTCTGCAATCACGGTGCCAGCATGGT
CAGGTTCTGGTGAGGGCCTCTTTCCTTCTCACTGTGTGCTCTTTCTTGTG
CATGGAGAGAGAGAGCATGAACAAGCCCTCTACTGTCCC
>Sequence 3370
AGGTACTTTGCCAAGCAGTAAAGGATCCAGGAGATAGCACTGGATGTGGT
GTCATGTCCTGCAAACATGAACGTTTTCACTTCAGCCTGGAGATCTGCTT
CAGAGAAATCTTTGGTGTTTTCGCTTTTGGCACTCAAAAGTATGTCCAGA
AAATCCCAGCGCCTTTTCTGAGTAGTATCTTGTTTTAGCTTATCCTTAAG
AGACTCCTTCCGGTCCTGGATTACTTTCTCTGTGAACTGATGAAGTTCTT
GGTTAAATTTAGAAAAGATTTGGCCTTGAGAGCTGAATTTGAAAACCAGG
TCGTTGTGATGTAGAAAATTGTTCATGCGCTGGTTGGAGATTTTGCTAAG
GTTGAACACTGCTTTCAGGTATGAGTCCAGGGTACCTGCCCGGGCGGCCG
CTC
>Sequence 3371
CCGGGCAGGTACCTGAGTAGTCTGTAACCAAAATAATTCATTTGGCACAT
TTAAGGCATCTGCCCAATAAAATACCTTGTGAATAAATTCTTTAGGAATT
ATGAATACTACAAGGAGAAAAATGACAGAAAGAAGCTGACTTGGGGGGAA
TGGGAGTTGTTTTTAATAACCTTTTTACCACCCTAGACTTCTGGCCAATC
AAATTGAATCATACCTTGGGAAGGGAAACTGCCTGATGCAGCAGCTGAGC
CAAAGTCAGCAAATCCTCCGAATAAATCAGCTGATCCTCCTAGAAGTTAA
GAAAAAGTTAAAACTGTTAGAACTAGAGAAAAATTGTAAATAATTTCTGG
AGGCCATTTTAGTGCCACACAAAAAGTACCT
>Sequence 3372
TTCTCTTACTGATAGTAGGATATTTCTGCTTTAGTTATTGTCACCTTAAA
TATATTTTCAATGTTGAAATCCTCACAGCATGTTTGATGAAATCTAGTTT
TCAAATTTTCTTAGGTATATTTCTGTCACGTTGGCATGATAACAAATGCA
ATAACCCAAAAGACCCCAAAAGCTAGTGTAATCCCTTTTGCAATCCAAGC
ATGAGGATTCATCTTCATGTTGACAGTGCGTGAATGTTCGGTAGGCTTTG
TCAAGCTTGCATACAATAAATTATATATGTCCCTTTTCTTTTAGGGTCTC
CTGTTAAAGATGGTCTTCTGAAGGCTAACTGCGGAATGAAAGTTTCTATT
CCAACTAAAGCCTTAAAATTGATGGACATGCAAACTTTCAAAGCAGAGCC
TCCCGAGAAGCC
>Sequence 3373
AGGTACTCACCCCAGTTCCTGCACCTTCTGTCTGCATACTCCTTGTCTCA
TAAGGGGTTTGAGTATGCACTCCAGCTGAACAGATGAGCCACACTCCTGT
CACATATCCTGCAAGGGGGTCAGGGAACTCTCCCATTTCAGCATGTGTGA

Table 4

TAAAAAATACATAATAATAAATGAATCTATCTATCAGATCTGTATAAAGA
AATAAATAAGAACTTTAGTCTTTCCTTTTATCCTGCTCTACTAGGAAAAA
TAAAAGTGAGCCAAACTTCAATATTAACAATCTCATTAATGGAATATATT
TTAAAAATGAATAATGTAGACATTATTTAACACTTTGACTATGGGTCAAC
TGTCCTTGGGCGTTTTAAACTAGTGGATCCCCCGGCTGGAGATTTAATTT
AAGCTTTTGATTCCGGCGACCTGAGG
>Sequence 3476
AGGTGCGCGGGTGTGCTGATGATTGATGTGGTGTGAGAACAAAGAAACAA
CAGAGAGTTTTCCCTACACACAGCCCAAAATACTATCTTAAAAAAATTTT
GAAGACAGGTTAACCAAAGGTTCTGGCCGCCATCCATGTCCAAAACAAAC
CAAATGTGTTGCCCGTAAAACATTTGCTCTTGAAAAGTAGTCCCTTAGCC
AGGCGCGGTGGCTCACACCTGCAATCCCAGCATTTGGAAGGCCGACGCAG
GTGGATCATCAGGTCAGGAGATCAAGACCATCCTGGCTAACGCAGTAAAA
TCCCATCTCTACTAAAAATACAAAAAGACTAGCCAGGCGTGGTGACACAC
GCCTGTAGTCCCAGCTACTCGGGAGGCTGACGCAGGAGAATCACTTGAAC
CCTGGAAGTGGAGGTTGCAGTGAGCGGTGATCGTGCCACCGCACTCCAGC
CTGGCAACAGAGCGAGACTCCGTCTCAAAAAAAACACCGCACCTAGAAGT
TCCTGCCCG
>Sequence 3477
AGGTACCAAATGCTAGTGAGGATGTGGAGCAACAGGAACTTTCATTCATT
GATGGTGGGAATGCAAAATGGAACAGCCACTTTGGAATACAATTTCTTAT
AAAACGAAACATACTCTTTTAAAAATAAATATACTCTTATCCTACGATCC
AGCAATCTCAATTTCCATGGTATTTATTCAAAGGAGTTGAAATTTATATC
CACACAAAAATCCACACATGGATGTTTGTGGCAGATTTATTCATAATTCT
AAGAAGTAGAAGCAACCAAGATGTCTTTCTGTAAGTGTCTTAATCCATTC
AACCTGATATACCAATGACAAAACTCATAGATTTGCTGGCATTTAAACAA
CATAAATTTATTGCTCACAGTCCTGGAGGCTGGAAGTCCAAGATCAAGTC
ATGGCACATTCTGCATCTGGTGAGGGCACTTCCTGGTTCATAGATGGCTC
CTTTTTGCTGCATCTAAGTGCACGAATCTCTTTCATGAGGTTCCACCCTC
ATGACCTCATCACCTCCCCAAAACTCCAACCTGGGCAACAGAGCAAGCCT
TTGTCTCGAAAAAAAAAAAAAAATGACAANAAAAGGTTCCTGCCCGGGCGG
CCGCTCGATGCCCATTTGATGCCTGGTTTATCACAAAAAAGACCCATGCA
CATCCCACTTTGCTCAGGGGAGGCGTTTCCCTAAGCAAGCTGGAAAAATA
GTGAAAACTGGTTTTTTTTTTTTTTTGGGGGCTTGGTTTTTACCCAAAAAT
TTTTCCTAGTATAAATGCCCTTTTTTTGAAA
>Sequence 3478
CGGGCAGGTACTGTCCAACTGGATGCTGCCCTGGTGGCTGAAGGCACACT
TCATGATGCTGTCCAGGGTCATCAGGGAGACATGTTGAAAGAGCTCCAGA
CGTGAGTTTTGGGCAATGTGTTCCTCCCATTTGTTCAGCATCATCCGAAC
ACTCTCAGACATCATGGTGATGAATATTTTCAGAATGCTGATGTTGAAGC
CAGGTTTCACAATCTGGCGGTGCTTTTTCCATTTAGAACCATCCAGGGTC
ACAAGTCCTCGACCAACCCAGGATTCAAGGATTTTGTGGCTAACAGCACT
TTTGGGATCTTGTCTTTTCAGGAGAATCTTGGCATAGTCTGGGTCATGGA
CACTGAAGAACATCGTAAAGGGTCCAACCCACAAGGGAACAGCACATGGG
TATTTTTCCATCAGCTTATGATACACCTCAAACTCCTTTACTGGGTAAAA
CTCCTTGTGGCCATAGAACCAGTGGGCAGGGGGTGCAGGAAACAGGTACA
GGGCTCTGATCATCCATCTTCTCCTCTGGTACCTT
>Sequence 3479
AGGTACCAGAAGGGCTCCCGCTCTCCTCTGCCCAAGAAGTGAGCGGTTCC
TCTGCCAGGACGGGGGGGAAGGAGGCCAACCTTGTAGGCTTCGTCAATGTT
GGCATTCACACAGTGCTGGATCATCTCCTTTACCAGCAGTGGGTGGGGCT
CGTCACAGACCTGGCCAAAGGGAAAGGAAAGGCGGTCAGGGGCTAGAAGG
GACAATGTGAGCGGGGAGGCCCCAACAGACCGACTTCCAGGCAATCTATG
GNTAAAACTTCTCAACCTGTTGTTTTCATCAAATTCTTCAGCAACCAGGA
TGAGTNTGGGGAATGCTCCTCTCACCATCAAAGAGAAGTGCTTTCGGCAC
CAAGAAGATAAATTTTCCAAACTTANTACAGATGAACAAGAAAGCCACCA

Table 4

AGCAGCCTCACGCGGGGCTTTCTGGTCTCGGCCGCAGAAGCGAGATGACG
AAGGGAACGTCATCGTTTGGAAAGCGTCGCAATAAGACGCACACGTTGTG
CCGCCGCTGTGGCTCTAAGGCCTACCACCTTCATAAGTCGACCTGTGGCA
AATGTGGCTACCCTGCCAAGCGCAAGAGAAAGTATAACTGGAGTGCCAAG
GCTAAAAGACTAAATACCACCGGAACTGGTCGAATGAGGCACCTAAAAAT
TGTATACCGCAAATTCAGGCATGGATTCCGTGAAGGAACAACAACCTAAA
CCCAAGAGGGCAGCTGTTGCAGCATCCAATTCATCTTAAGAATGTCAACG
ATTAGTCATGCAATAAATGTTCTGGTTTTGAAAAAAAAAGCATTTAAAAA
AGGTACCTGCCCG

>Sequence 4134

TCCGCCGCCCACACAATTGAGGATGTGTGTATCTATATGAATCTNAATTT
CTATCTCTCACTCACACTATATTGATTGTATATGTGTAGATTGAGAGGAG
AATATTACCCATAAGTGAATGATTCTCTTCCCCCTGCTTGAGAGGAGCCC
GGAGATGAACT

>Sequence 4135

AGGTACCCTGAGCATCAGCAGAGATGCCGAGATGAAATCAGGGAACTCCT
AGGGGATGGGTCTTCTATTACCTGGGAACACCTGAGCCAGATGCCTTACA
CCACGATGTGCATCAAGGAATGCCTCCGCCTCTACGCACCGGTAGTAAAC
ATATCCCGGTTACTCGACAAACCCATTACTTTTCCAGATGGACGCTCCTT
ACCTGCAGGAATAACTGTGTTTATCAATATTTGGGCTCTTCACCACAACC
CCTATTTCTGGGAAGACCCTCAGGTCTTTAACCCCTTGAGATTCTCCAGG
GAAAATTCTGAAAAAATACATCCCTATGCCTTCATACCATTCTCAGCTGG
ATTAAGGAACTGCATTGGGCAGCAATTTTGCCATAATTGAGTGTAAAGTG
GCAGTGGCATTAACTCTGCTCCGCTTTCAAGCTGGCTCCAGACCACTCAA
GGCCTTCCCAGCCTGTTCGTCAAGTTGTCCTCAAAGTCCCAAAATGGAAT
CCAATGTGTTTGCAAAAAAAAAAGTTTGCTAATTTTAAGTCCCCTTTCGTT
AAGAATTAATGAGACCATTTTTCTACCAAGGGAGAACCAAAGGGTTAAAT
TTATTCCAAATTTTTGTTATGGGTGTTGGCAAATTATTTACTTTGGATAC
CTTTGACGGGTTTTACATCATTAACCAGAATTTTAAATTTTTTGCTGGTT
GGGGAAAACCCCAAAAACCCCGAAAAAATAAGAGTGTTTTCTTGGGAGGG
GAAATTG

>Sequence 4136

TCCACCGCGGTGGCGGCCGAGGTACCTGACCAGACACACCCAAATGCATG
TCATACTGCACAAATTACACATTGCAAAGTGTGTCCTCTGCACTGTTATG
GAAAACAATTTGCAGAATACTTCCTTTTCCCTTCTGTTGTGCCCAACTCT
TTATTATAGGATAGGGATTTAAGGTCAGGGTTACATGGAAGCAGGGTTCC
CTGGGGTGTTGTATCAGCCTGATGCCGCCCCACCTCAGCTGCCCCTCCAC
TTTCATCTCCCATTCTTCTTTATGAGAGCAGAGAGGAGAAGGGAAAGACG
TGGCAGAGGAATAAGCAGAGATGGTTACCCTGAGACAGGACTTTGGTCCA
CATCATTTACTTGAGAGGTGATCCCAGGAACCATGGGGACAGTGACACAA
GGAAGACAGGAAGCCAATGAAGGGTATGATGACGAGTGGGTGCCTGGCAG
TGAGTGGACAGGCAAGGCTCAGCCTGCTGGGGCCCCCTGAGAGACTGCAG
AACTTCTCCCCAGATTGTCCCAGTGAAGAGCAAGGGAGTGGGGTTACTGA
TCCACCAACTTGCCTCTGTTGGTTGAGAGTTGCTTTTAACCCAAGGCATC
AACTCCCTGGTACCTGCCCCGGCCGCCGCTTTAAAACTAAGGGAATCCCC
CGGCTTGAGGAATTTCATTTAAAGCTTATTGATCCCGTCCACCTCGAGGG
GGGCCCCGGCCCCAACTTTGTTCCCCTTAGGGAGGGGTAATGGGCCCTTG
GGGAAAAATGGGAAAAAATTTTTCCTGGG

>Sequence 4137

AGGTACCTTAATACACTCCAAACCTCAGACTGGTGATATTGCATTACTTT
CTATAAGATACACACTTATATTTTGCTTAAAGAAGCATTTGCAGCTTTAA
CCCATTTAATGTAATATATAAGGCAGATGGTTCTTGACAACTATAATCTT
GGAAGAAAACTGAAAAAATCAATTAGGTGTTTGGCAGCACTCAAAAAACA
GCCTAATAAGCTCATTGCCATGCAAATCACCTACAATGTTAAATATATAA
AACTGCTGTATTAAGCTCGGAGAAGGAGGAAAATTTCCTCTTCACTCATG
TGACCGGAGACTAAAAGGAAGGCTCCAGGGGTGCTGCTGGCTGGGAAGAG